36/PRTS

09/720086

JC01 Rec'd PCT/PTO 20 DEC 2000

WO 99/67397

PCT/US99/14373

De novo DNA Cytosine Methyltransferase Genes, Polypeptides and Uses Thereof

Background of the Invention

Field of the Invention

The present invention relates generally to the fields of molecular biology, developmental biology, cancer biology and medical therapeutics. Specifically, the present invention relates to novel DNA cytosine methyltransferases. More specifically, isolated nucleic acid molecules are provided encoding mouse Dnmt3a and Dnmt3b and human DNMT3A and DNMT3B *de novo* DNA cytosine methyltransferase genes. Dnmt3a and Dnmt3b mouse and DNMT3A and DNMT3B human polypeptides are also provided, as are vectors, host cells and recombinant methods for producing the same. The invention further relates to an *in vitro* method for cytosine C5 methylation. Also provided is a diagnostic method for neoplastic disorders, and methods of gene therapy using the polynucleotides of the invention.

Related Art

Methylation at the C-5 position of cytosine predominantly in CpG dinucleotides is the major form of DNA modification in vertebrate and invertebrate animals, plants, and fungi. Two distinctive enzymatic activities have been shown to be present in these organisms. The *de novo* DNA cytosine methyltransferase, whose expression is tightly regulated in development, methylates unmodified CpG sites to establish tissue or gene-specific methylation patterns. The maintenance methyltransferase transfers a methyl group to cytosine in hemi-methylated CpG sites in newly replicated DNA, thus functioning to maintain clonal inheritance of the existing methylation patterns.

De novo methylation of genomic DNA is a developmentally regulated process (Jähaner, D. and Jaenish, R., "DNA Methylation in Early Mammalian Development," In *DNA Methylation: Biochemistry and Biological Significance*, Razin, A. *et al., eds.*, Springer-Verlag (1984) pp. 189-219 and Razin, A., and Cedar, H., "DNA Methylation and Embryogenesis," in *DNA Methylation: Molecular Biology and Biological Significance*, Jost., J. P. *et al., eds.*, Birkhäuser Verlag, Basel, Switzerland (1993) pp. 343-357). It plays a pivotal role in the establishment of parental-specific methylation patterns of imprinted genes (Chaillet, J. R. *et al., Cell 66*:77-83 (1991); Stöger, R. *et al., Cell* 73:61-71 (1993); Brandeis, M. *et al., EMBO J. 12*:3669-3677 (1993); Tremblay, K. D. *et al., Nature Genet. 9*:407-413 (1995); and Tucker, K. L. *et al., Genes Dev. 10*:1008-1020 (1996)), and in the regulation of X chromosome inactivation in mammals (Brockdoff, N. "Convergent Themes in X Chromosome Inactivation and Autosomal Imprinting," in *Genomic Imprinting: Frontiers in Molecular Biology*, Reik, W. and Sorani, A. *eds.*, IRL Press Oxford (1997) pp. 191-210; Ariel, M. *et al., Nature Genet. 9*:312-315 (1995); and Zucotti, M. and Monk, M. *Nature Genet. 9*:316-320 (1995)).

Thus, C5 methylation is a tightly regulated biological process important in the control of gene regulation. Additionally, aberrant *de novo* methylation can lead to undesirable consequences. For example, *de novo* methylation of growth regulatory genes in somatic tissues is associated with tumorigenesis in humans (Laird, P. W. and Jaenisch, R. *Ann. Rev. Genet. 30*:441-464 (1996); Baylin, S. B. *et al., Adv. Cancer. Res. 72*:141-196 (1998); and Jones, P. A. and Gonzalgo, M. L. *Proc. Natl. Acad. Sci. USA 94*:2103-2105 (1997)).

The gene encoding the major maintenance methyltransferase, *Dnmt*1, was first cloned in mice (Bestor, T. H. *et al., J. Mol. Biol. 203*:971-983 (1988), and the homologous genes were subsequently cloned from a number of organisms, including *Arabidoposis*, sea urchin, chick, and human. *Dnmt*1 is expressed ubiquitously in human and mouse tissues. Targeted disruption of *Dnmt*1 results in a genome-wide loss of cytosine methylation and embryonic lethality (Li *et al.*,

1992). Interestingly, *Dnmt*1 is dispensable for the survival and growth of the embryonic stem cells, but appears to be required for the proliferation of differentiated somatic cells (Lei *et al.*, 1996). Although it has been shown that the enzyme encoded by *Dnmt*1 can methylate DNA *de novo in vitro* (Bestor, 1992), there is no evidence that *Dnmt*1 is directly involved in *de novo* methylation in normal development. *Dnmt*1 appears to function primarily as a maintenance methyltransferase because of its strong preference for hemi-methylated DNA and direct association with newly replicated DNA (Leonhardt, H. *et al., Cell 71*:865-873 (1992)). Additionally, ES cells homozygous for a null mutation of *Dnmt*1 can methylate newly integrated retroviral DNA, suggesting that *Dnmt*1 is not required for *de novo* methylation and an independently encoded *de novo* DNA cytosine methyltransferase is present in mammalian cells (Lei *et al.*, 1996).

Various methods of disrupting *Dnmt*1 protein activity are known to those skilled in the art. For example, see PCT Publication No. WO92/06985, wherein mechanism based inhibitors are discussed. Applications involving antisense technology are also known; U.S. Patent No. 5578716 discloses the use of antisense oligonucleotides to inhibit *Dnmt*1 activity, and Szyf *et al., J. Biol. Chem.* 267: 12831-12836, 1992, demonstrates that myogenic differentiation can be affected through the antisense inhibition of *Dnmt*1 protein activity.

Thus, while there is a significant amount of knowledge in the art regarding the maintenance C5 methyltransferase (*Dnmt*1), there is no information regarding nucleic acid or protein structure and expression or enzymatic properties of the *de novo* C5 methyltransferase in mammals.

Summary of the Invention

A first aspect of the invention provides novel *de novo* DNA cytosine methyltransferase nucleic acids and polypeptides that are not available in the art. A second aspect of the invention relates to *de novo* DNA cytosine methyltransferase recombinant materials and methods for their production. A

third aspect of the invention relates to the production of recombinant *de novo* DNA cytosine methyltransferase polypeptides. A fourth aspect of the invention relates to methods for using such *de novo* DNA cytosine methyltransferase polypeptides and polynucleotides. Such uses include the treatment of neoplastic disorders, among others. Yet another aspect of the invention relates to diagnostic assays for the detection of diseases associated with inappropriate *de novo* DNA cytosine methyltransferase activity or levels and mutations in *de novo* DNA cytosine methyltransferases that might lead to neoplastic disorders.

Brief Description of the Figures

Figure 1A-1D shows the nucleotide sequences of mouse Dnmt3a and Dnmt3b and human DNMT3A and DNMT3B genes, respectively.

Figure 2A-2D shows the deduced amino acid sequence of mouse Dnmt3a and Dnmt3b and human DNMT3A and DNMT3B genes, respectively. Sequences are presented in single letter amino acid code.

Figure 3A shows a comparison of mouse Dnmt3a and Dnmt3b amino acid sequences, and Figure 3B presents a comparison of the protein sequences of human DNMT3A and DNMT3B1.

Figure 4A presents a schematic comparison of mouse *Dnmt1, Dnmt2,* Dnmt3a and Dnmt3b protein structures. Figure 4B presents a schematic of the DNMT3A, DNMT3B and zebrafish Zmt3 proteins. Figure 4C and 4D present a schematic of the human DNMT3B gene organization and exon/intron junction sequences.

Figure 5A presents a comparison of highly conserved protein structural motifs for eukaryotic and prokaryotic C5 methyltransferase. Figure 5B presents a sequence alignment of the C-rich domain of vertebrate DNMT3 proteins and the X-lined ATRX gene. Figure 5C presents a non-rooted phylogenic tree of methyltransferase proteins..

Figure 6A-6C demonstrates the expression of Dnmt3a and Dnmt3b in mouse adult tissues, embryos, and ES cells by northern blot.

Figure 7A-7D demonstrates *in vitro* methyltransferase activities of mouse Dnmt3a and Dnmt3b proteins.

Figure 8 demonstrates *in vitro* analysis of *de novo* and maintenance activities of Dnmt3a, Dnmt3b1 and Dnmt3b2 proteins.

Figure 9 presents Northern blot expression analysis of DNMT3A and DNMT3B.

Figure 10 presents DNMT3 Northern Blot expression analysis of DNMT3A and DNMT3B in human tumor cell lines.

Detailed Description of the Preferred Embodiments

Definitions

In the description that follows, a number of terms used in recombinant DNA technology are utilized extensively. In order to provide a clear and consistent understanding of the specification and claims, including the scope to be given such terms, the following definitions are provided.

Cloning vector: A plasmid or phage DNA or other DNA sequence which is able to replicate autonomously in a host cell, and which is characterized by one or a small number of restriction endonuclease recognition sites at which such DNA sequences may be cut in a determinable fashion without loss of an essential biological function of the vector, and into which a DNA fragment may be spliced in order to bring about its replication and cloning. The cloning vector may further contain a marker suitable for use in the identification of cells transformed with the cloning vector. Markers, for example, provide tetracycline resistance or ampicillin resistance.

Expression vector: A vector similar to a cloning vector but which is capable of enhancing the expression of a gene which has been cloned into it, after

transformation into a host. The cloned gene is usually placed under the control of (i.e., operably linked to) certain control sequences such as promoter sequences. Promoter sequences may be either constitutive or inducible.

Recombinant Host: According to the invention, a recombinant host may be any prokaryotic or eukaryotic host cell which contains the desired cloned genes on an expression vector or cloning vector. This term is also meant to include those prokaryotic or eukaryotic cells that have been genetically engineered to contain the desired gene(s) in the chromosome or genome of that organism. For examples of such hosts, *see* Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*, Second Edition, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York (1989). Preferred recombinant hosts are eukaryotic cells transformed with the DNA construct of the invention. More specifically, mammalian cells are preferred.

Recombinant vector: Any cloning vector or expression vector which contains the desired cloned gene(s).

Host Animal: Transgenic animals, all of whose germ and somatic cells contain the DNA construct of the invention. Such transgenic animals are in general vertebrates. Preferred Host Animals are mammals such as non-human primates, mice, sheep, pigs, cattle, goats, guinea pigs, rodents, e.g. rats, and the like. The term Host Animal also includes animals in all stages of development, including embryonic and fetal stages.

Promoter: A DNA sequence generally described as the 5′ region of a gene, located proximal to the start codon. The transcription of an adjacent gene(s) is initiated at the promoter region. If a promoter is an inducible promoter, then the rate of transcription increases in response to an inducing agent. In contrast, the rate of transcription is not regulated by an inducing agent if the promoter is a constitutive promoter. According to the invention, preferred promoters are heterologous to the *de novo* DNA cytosine methyltransferase genes, that is, the promoters do not drive expression of the gene in a mouse or human. Such promoters include the CMV promoter (InVitrogen, San Diego, CA), the

SV40, MMTV, and hMTIIa promoters (U.S. 5,457,034), the HSV-1 4/5 promoter (U.S. 5,501,979), and the early intermediate HCMV promoter (WO92/17581). In one emdodiment, it is preferred that the promoter is tissue-specific, that is, it is induced selectively in a specific tissue. Also, tissue-specific enhancer elements may be employed. Additionally, such promoters may include tissue and cell-specific promoters of an organism.

Gene: A DNA sequence that contains information needed for expressing a polypeptide or protein.

Structural gene: A DNA sequence that is transcribed into messenger RNA (mRNA) that is then translated into a sequence of amino acids characteristic of a specific polypeptide.

Complementary DNA (cDNA): A "complementary DNA," or "cDNA" gene includes recombinant genes synthesized by reverse transcription of mRNA and from which intervening sequences (introns) have been removed.

Expression: Expression is the process by which a polypeptide is produced from a structural gene. The process involves transcription of the gene into mRNA and the translation of such mRNA into polypeptide(s).

Homologous/Nonhomologous: Two nucleic acid molecules are considered to be "homologous" if their nucleotide sequences share a similarity of greater than 40%, as determined by HASH-coding algorithms (Wilber, W.J. and Lipman, D.J., *Proc. Natl. Acad. Sci. 80*:726-730 (1983)). Two nucleic acid molecules are considered to be "nonhomologous" if their nucleotide sequences share a similarity of less than 40%.

Polynucleotide: This term generally refers to any polyribonucleotide or polydeoxyribonucleotide, which may be unmodified RNA or DNA or modified RNA or DNA. "Polynucleotides" include, without limitation single- and double-stranded DNA, DNA that is a mixture of single- and double-stranded regions, single- and double-stranded RNA, and RNA that is mixture of single- and double-stranded regions, hybrid molecules comprising DNA and RNA that may be single-stranded or, more typically, double-stranded or a mixture of single- and

double-stranded regions. In addition, "polynucleotide" refers to triple-stranded regions comprising RNA or DNA or both RNA and DNA. The term polynucleotide also includes DNAs or RNAs containing one or more modified bases and DNAs or RNAs with backbones modified for stability or for other reasons. "Modified" bases include, for example, tritylated bases and unusual bases such as inosine. A variety of modifications have been made to DNA and RNA; thus, "polynucleotide" embraces chemically, enzymatically or metabolically modified forms of polynucleotides as typically found in nature, as well as the chemical forms of DNA and RNA characteristic of viruses and cells. "Polynucleotide" also embraces relatively short polynucleotides, often referred to as oligonucleotides.

Polypeptide: This term refers to any peptide or protein comprising two or more amino acids joined to each other by peptide bonds or modified peptide bonds, *i.e.*, peptide isosteres. "Polypeptide" refers to both short chains, commonly referred to as peptides, oligopeptides or oligomers, and to longer chains, generally referred to as proteins. Polypeptides may contain amino acids other than the 20 gene-encoded amino acids. "Polypeptides" include amino acid sequences modified either by natural processes, such as post-translational processing, or by chemical modification techniques which are well known in the art. Such modifications are well described in basic texts and in more detailed monographs, as well as in a voluminous research literature. Modifications can occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. It will be appreciated that the same type of modification may be present in the same or varying degrees at several sites in a given polypeptide. Also, a given polypeptide may contain many types of modifications. Polypeptides may be branched as a result of ubiquitination, and they may be cyclic, with or without branching. Cyclic, branched and branched cyclic polypeptides may result from post-translation natural processes or may be made by synthetic methods. Modifications include acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of

flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphotidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent cross-links, formation of cystine, formation of pyroglutamate, formylation, gamma-carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination. See, for instance, *Proteins-Structure and Molecular Properties*, 2nd Ed., T. E. Creighton, W. H. Freeman and Company, New York, 1993 and Wold, F., Posttranslational Protein Modifications: Perspectives and Prospects, pgs. 1-12 in *Posttranslational Covalent Modification of Proteins*, B. C. Johnson, Ed., Academic Press, New York, 1983; Seifter *et al.*, "Analysis for protein modifications and nonprotein cofactors", *Methods in Enzymol. 182*:626-646 (1990) and Rattan *et al.*, "Protein Synthesis: Posttranslational Modifications and Aging", *Ann NY Acad Sci 663*:48-62 (1992).

Variant: The term used herein is a polynucleotide or polypeptide that differs from a reference polynucleotide or polypeptide respectively, but retains essential properties. A typical variant of a polynucleotide differs in nucleotide sequence from another, reference polynucleotide. Changes in the nucleotide sequence of the variant may or may not alter the amino acid sequence of a polypeptide encoded by the reference polynucleotide. Nucleotide changes may result in amino acid substitutions, additions, deletions, fusions and truncations in the polypeptide encoded by the reference sequence, as discussed below. A typical variant of a polypeptide differs in amino acid sequence from another, reference polypeptide. Generally, differences are limited so that the sequences of the reference polypeptide and the variant are closely similar overall and, in many regions, identical. A variant and reference polypeptide may differ in amino acid sequence by one or more substitutions, additions, deletions in any combination.

A substituted or inserted amino acid residue may or may not be one encoded by the genetic code. A variant of a polynucleotide or polypeptide may be a naturally occurring such as an allelic variant, or it may be a variant that is not known to occur naturally. Non-naturally occurring variants of polynucleotides and polypeptides may be made by mutagenesis techniques or by direct synthesis.

Identity: This term refers to a measure of the identity of nucleotide sequences or amino acid sequences. In general, the sequences are aligned so that the highest order match is obtained. "Identity" *per se* has an art-recognized meaning and can be calculated using published techniques. (See, *e.g.*: *Computational Molecular Biology*, Lesk, A.M., ed., Oxford University Press, New York, 1988; *Biocomputing: Informatics and Genome Projects*, Smith, D.W., ed., Academic Press, New York, 1993; *Computer Analysis of Sequence Data, Part I*, Griffin, A.M., and Griffin, H.G., eds., Humana Press, New Jersey, 1994; *Sequence Analysis in Molecular Biology*, von Heinje, G., Academic Press, 1987; and *Sequence Analysis Primer*, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991). While there exist a number of methods to measure identity between two polynucleotide or polypeptide sequences, the term "identity" is well known to skilled artisans (Carillo, H. & Lipton, D., *SIAM J Applied Math 48*:1073 (1988)). Methods commonly employed to determine identity or similarity between two sequences include, but are not limited to, those disclosed in Guide to Huge Computers, Martin J. Bishop, ed., Academic Press, San Diego, 1994, and Carillo, H. & Lipton, D., *SIAM J Applied Math 48*:1073 (1988). Methods to determine identity and similarity are codified in computer programs. Preferred computer program methods to determine identity and similarity between two sequences include, but are not limited to, GCS program package (Devereux, J., *et al.*, *Nucleic Acids Research 12(I)*:387 (1984)), BLASTP, BLASTN, FASTA (Atschul, S.F., *et al.*, *J Mol. Biol 215*:403 (1990)).

Therefore, as used herein, the term "identity" represents a comparison between a test and reference polynucleotide. More specifically, reference polynucleotides are identified in this invention as SEQ ID Nos: 1, 2, 3 and 4, and

a test polynucleotide is defined as any polynucleotide that is 90% or more identical to a reference polynucleotide. As used herein, the term "90% or more" refers to percent identities from 90 to 99.99 relative to the reference polynucleotide. Identity at a level of 90% or more is indicative of the fact that, assuming for exemplification purposes a test and reference polynucleotide length of 100 nucleotides, that no more than 10% (i.e., 10 out of 100) nucleotides in the test polynucleotide differ from that of the reference polynucleotide. Such differences may be represented as point mutations randomly distributed over the entire length of the sequence or they may be clustered in one or more locations of varying length up to the maximum allowable 10 nucleotide difference. Differences are defined as nucleotide substitutions, deletions or additions of sequence. These differences may be located at any position in the sequence, including but not limited to the 5' end, 3' end, coding and non coding sequences.

Fragment: A "fragment" of a molecule such as *de novo* DNA cytosine methyltransferases is meant to refer to any polypeptide subset of that molecule.

Functional Derivative: The term "functional derivatives" is intended to include the "variants," "analogues," or "chemical derivatives" of the molecule. A "variant" of a molecule such as *de novo* DNA cytosine methyltransferases is meant to refer to a naturally occurring molecule substantially similar to either the entire molecule, or a fragment thereof. An "analogue" of a molecule such as *de novo* DNA cytosine methyltransferases is meant to refer to a non-natural molecule substantially similar to either the entire molecule or a fragment thereof.

A molecule is said to be "substantially similar" to another molecule if the sequence of amino acids in both molecules is substantially the same, and if both molecules possess a similar biological activity. Thus, provided that two molecules possess a similar activity, they are considered variants as that term is used herein even if one of the molecules contains additional amino acid residues not found in the other, or if the sequence of amino acid residues is not identical.

As used herein, a molecule is said to be a "chemical derivative" of another molecule when it contains additional chemical moieties not normally a part of the

molecule. Such moieties may improve the molecule's solubility, absorption, biological half-life, etc. The moieties may alternatively decrease the toxicity of the molecule, eliminate or attenuate any undesirable side effect of the molecule, etc. Examples of moieties capable of mediating such effects are disclosed in *Remington's Pharmaceutical Sciences* (1980) and will be apparent to those of ordinary skill in the art.

Protein Activity or ***Biological Activity of the Protein:*** These expressions refer to the metabolic or physiologic function of *de novo* DNA cytosine methyltransferase protein including similar activities or improved activities or these activities with decreased undesirable side-effects. Also included are antigenic and immunogenic activities of said *de novo* DNA cytosine methyltransferase protein. Among the physiological or metabolic activities of said protein is the transfer of a methyl group to the cytosine C5 position of duplex DNA. Such DNA may completely lack any methylation of may be hemimethylated. As demonstrated in Example 8, *de novo* DNA cytosine methyltransferases methylate C5 in cytosine moieties in nonmethylated DNA.

De novo DNA Cytosine Methyltransferases Polynucleotides: This term refers to a polynucleotide containing a nucleotide sequence which encodes a *de novo* DNA cytosine methyltransferase polypeptide or fragment thereof or that encodes a *de novo* DNA cytosine methyltransferase polypeptide or fragment wherein said nucleotide sequence has at least 90% identity to a nucleotide sequence encoding the polypeptide of SEQ ID Nos: 5, 6, 7 or 8, or a corresponding fragment thereof, or which has sufficient identity to a nucleotide sequence contained in SEQ ID NO:1, 2, 3 or 4.

De novo DNA Cytosine Methyltransferases Polypeptides: This term refers to polypeptides with amino acid sequences sufficiently similar to the *de novo* DNA cytosine methyltransferase protein sequence in SEQ ID NO:5, 6, 7 or 8 and that at least one biological activity of the protein is exhibited.

Antibodies: As used herein includes polyclonal and monoclonal antibodies, chimeric, single chain, and humanized antibodies, as well as Fab

fragments, including the products of an Fab or other immunoglobulin expression library.

Substantially pure: As used herein means that the desired purified protein is essentially free from contaminating cellular components, said components being associated with the desired protein in nature, as evidenced by a single band following polyacrylamide-sodium dodecyl sulfate gel electrophoresis. Contaminating cellular components may include, but are not limited to, proteinaceous, carbohydrate, or lipid impurities.

The term "substantially pure" is further meant to describe a molecule which is homogeneous by one or more purity or homogeneity characteristics used by those of skill in the art. For example, a substantially pure *de novo* DNA cytosine methyltransferases will show constant and reproducible characteristics within standard experimental deviations for parameters such as the following: molecular weight, chromatographic migration, amino acid composition, amino acid sequence, blocked or unblocked N-terminus, HPLC elution profile, biological activity, and other such parameters. The term, however, is not meant to exclude artificial or synthetic mixtures of the factor with other compounds. In addition, the term is not meant to exclude *de novo* DNA cytosine methyltransferase fusion proteins isolated from a recombinant host.

Isolated: A term meaning altered "by the hand of man" from the natural state. If an "isolated" composition or substance occurs in nature, it has been changed or removed from its original environment, or both. For example, a polynucleotide or a polypeptide naturally present in a living animal is not "isolated," but the same polynucleotide or polypeptide separated from the coexisting materials of its natural state is "isolated", as the term is employed herein. Thus, a polypeptide or polynucleotide produced and/or contained within a recombinant host cell is considered isolated for purposes of the present invention. Also intended as an "isolated polypeptide" or an "isolated polynucleotide" are polypeptides or polynucleotides that have been purified, partially or substantially, from a recombinant host cell or from a native source.

For example, a recombinantly produced version of a *de novo* DNA cytosine methyltransferase polypeptide can be substantially purified by the one-step method described in Smith and Johnson, *Gene 67:*31-40 (1988).

Neoplastic disorder: This term refers to a disease state which is related to the hyperproliferation of cells. Neoplastic disorders include, but are not limited to, carcinomas, sarcomas and leukemias.

Gene Therapy: A means of therapy directed to altering the normal pattern of gene expression of an organism. Generally, a recombinant polynucleotide is introduced into cells or tissues of the organism to effect a change in gene expression.

Antisense RNA gene/Antisense RNA. In eukaryotes, mRNA is transcribed by RNA polymerase II. However, it is also known that one may construct a gene containing a RNA polymerase II template wherein a RNA sequence is transcribed which has a sequence complementary to that of a specific mRNA but is not normally translated. Such a gene construct is herein termed an "antisense RNA gene" and such a RNA transcript is termed an "antisense RNA." Antisense RNAs are not normally translatable due to the presence of translation stop codons in the antisense RNA sequence.

Antisense oligonucleotide: A DNA or RNA molecule or a derivative of a DNA or RNA molecule containing a nucleotide sequence which is complementary to that of a specific mRNA. An antisense oligonucleotide binds to the complementary sequence in a specific mRNA and inhibits translation of the mRNA. There are many known derivatives of such DNA and RNA molecules. See, for example, U.S. Patent Nos. 5,602,240, 5,596,091, 5,506,212, 5,521,302, 5,541,307, 5,510,476, 5,514,787, 5,543,507, 5,512,438, 5,510,239, 5,514,577, 5,519,134, 5,554,746, 5,276,019, 5,286,717, 5,264,423, as well as WO96/35706, WO96/32474, WO96/29337 (thiono triester modified antisense oligodeoxynucleotide phosphorothioates), WO94/17093 (oligonucleotide alkylphosphonates and alkylphosphothioates), WO94/08004 (oligonucleotide phosphothioates, methyl phosphates, phosphoramidates, dithioates, bridged

phosphorothioates, bridge phosphoramidates, sulfones, sulfates, ketos, phosphate esters and phosphorobutylamines (van der Krol *et al.*, *Biotech.* *6*:958-976 (1988); Uhlmann *et al.*, *Chem. Rev.* *90*:542-585 (1990)), WO94/02499 (oligonucleotide alkylphosphonothioates and arylphosphonothioates), and WO92/20697 (3'-end capped oligonucleotides). Particular *de novo* DNA cytosine methyltransferase antisense oligonucleotides of the present invention include derivatives such as S-oligonucleotides (phosphorothioate derivatives or S-oligos, *see*, Jack Cohen, *Oligodeoxynucleotides, Antisense Inhibitors of Gene Expression*, CRC Press (1989)). S-oligos (nucleoside phosphorothioates) are isoelectronic analogs of an oligonucleotide (O-oligo) in which a nonbridging oxygen atom of the phosphate group is replaced by a sulfur atom. The S-oligos of the present invention may be prepared by treatment of the corresponding O-oligos with 3*H*-1,2-benzodithiol-3-one-1,1-dioxide which is a sulfur transfer reagent. *See* Iyer *et al.*, *J. Org. Chem.* *55*:4693-4698 (1990); and Iyer *et al.*, *J. Am. Chem. Soc.* *112*:1253-1254 (1990).

Antisense Therapy: A method of treatment wherein antisense oligonucleotides are administered to a patient in order to inhibit the expression of the corresponding protein.

I. Deposited Material

The invention relates to polynucleotides encoding and polypeptides of novel *de novo* DNA cytosine methyltransferase proteins. The invention relates especially to *de novo* DNA cytosine methyltransferase mouse Dnmt3a and Dnmt3b cDNAs and the human DNMT3A and *DNMT3B* cDNAs set out in SEQ ID NOs:1, 2, 3 and 4, respectively. The invention also relates to mouse Dnmt3a and Dnmt3b nd human DNMT3A and DNMTB *e novo* DNA cytosine methyltransferase polypeptides set out in SEQ ID NOs:5, 6, 7, and 8, respectively. The invention further relates to the *de novo* DNA cytosine methyltransferase nucleotide sequences of the mouse Dnmt3a cDNA (plasmid pMT3a) and Dnmt3b cDNA (plasmid pMT3b) and the human *DNMTα* cDNA (plasmid pMT3A) in

ATCC Deposit Nos.209933, 209934, and 98809, respectively, and the amino acid sequences encoded therein.

The nucleotide sequence of the human *DNMT3B* cDNA identified in SEQ ID NO:4 is available in a clone (ATCC Deposit No. 326637) independently deposited by the I.M.A.G.E. Consortium. The invention relates to the *de novo* DNA cytosine methyltransferase polypeptide encoded therein.

Clones containing mouse Dnmt3a and Dnmt3b cDNAs were deposited with the American Type Culture Collection (ATCC), Manassas, Virginia 20110-2209, USA, on June 16, 1998, and assigned ATCC Deposit Nos. 209933 and 209934, respectively. The human DNMT3A cDNA was deposited with the ATCC on July 10, 1998, and assigned ATCC Deposit No. 98809.

While the ATCC deposits are believed to contain the *de novo* DNA cytosine methyltransferase cDNA sequences shown in SEQ ID NOs:1, 2, 3, and 4, the nucleotide sequences of the polynucleotide contained in the deposited material, as well as the amino acid sequence of the polypeptide encoded thereby, are controlling in the event of any conflict with any description of sequences herein.

The deposits for mouse Dnmt3a and Dnmt3b cDNAs and the human DNMT3A cDNA were made under the terms of the Budapest Treaty on the international recognition of the deposit of micro-organisms for purposes of patent procedure. The deposits are provided merely as a convenience for those of skill in the art and are not an admission that a deposit is required for enablement, such as that required under 35 U.S.C. § 112.

II. Polynucleotides of the Invention

Another aspect of the invention relates to isolated polynucleotides, and polynucleotides closely related thereto, which encode the *de novo* DNA cytosine methyltransferase polypeptides. As shown by the results presented in Figure 5, sequencing of the cDNAs contained in the deposited clones encoding mouse and

human *de novo* DNA cytosine methyltransferases confirms that the *de novo* DNA cytosine methyltransferase proteins of the invention are structurally related to other proteins of the DNA methyltransferase family.

The polynucleotides of the present invention encoding *de novo* DNA cytosine methyltransferase proteins may be obtained using standard cloning and screening procedures as described in Example 1. Polynucleotides of the invention can also be obtained from natural sources such as genomic DNA libraries or can be synthesized using well known and commercially available techniques.

Among particularly preferred embodiments of the invention are polynucleotides encoding *de novo* DNA cytosine methyltransferase polypeptides having the amino acid sequence set out in SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, or SEQ ID NO:8, and variants thereof.

A particular nucleotide sequence encoding a *de novo* DNA cytosine methyltransferase polypeptide may be identical over its entire length to the coding sequence in SEQ ID NOs:1, 2, or 3. Alternatively, a particular nucleotide sequence encoding a *de novo* DNA cytosine methyltransferase polypeptide may be an alternate form of SEQ ID NOs:1, 2, 3 and 4 due to degeneracy in the genetic code or variation in codon usage encoding the polypeptides of SEQ ID NOs:5, 6, 7, or 8. Preferably, the polynucleotides of the invention contain a nucleotide sequence that is highly identical, at least 90% identical, with a nucleotide sequence encoding a *de novo* DNA cytosine methyltransferase polypeptide or at least 90% identical with the encoding nucleotide sequence set forth in SEQ ID NOs:1, 2, or 3. Polynucleotides of the invention may be 90 to 99% identical to the nucleotides sequence set forth in SEQ ID NO:4.

When a polynucleotide of the invention is used for the recombinant production of a *de novo* DNA cytosine methyltransferase polypeptide, the polynucleotide may include the coding sequence for the full-length polypeptide or a fragment thereof, by itself; the coding sequence for the full-length polypeptide or fragment in reading frame with other coding sequences, such as those encoding a leader or secretory sequence, a pre-, or pro or prepro-protein

sequence, or other fusion peptide portions. For example, a marker sequence that facilitates purification of the fused polypeptide can be encoded. In certain preferred embodiments of this aspect of the invention, the marker sequence is a hexa-histidine peptide, as provided in the pQE vector (Qiagen, Inc.) and described in Gentz *et al.*, *Proc Natl Acad Sci USA 86*:821-824 (1989), or it may be the HA tag, which corresponds to an epitope derived from the influenza hemagglutinin protein (Wilson, I., *et al., Cell 37*:767, 1984). The polynucleotide may also contain non-coding 5′ and 3′ sequences, such as transcribed, non-translated sequences, splicing and polyadenylation signals, ribosome binding sites and sequences that stabilize mRNA.

Embodiments of the invention include isolated nucleic acid molecules comprising a polynucleotide having a nucleotide sequence at least 90% identical, and more preferably at least 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% identical to (a) a nucleotide sequence encoding a *de novo* DNA cytosine methyltransferase polypeptide having the amino acid sequence in SEQ ID NO:1, SEQ ID NO:2, or SEQ ID NO:3; (b) a nucleotide sequence encoding a *de novo* DNA cytosine methyltransferase polypeptide having the amino acid sequence encoded by the cDNA clone contained in ATCC Deposit No. 209933, ATCC Deposit No. 209934, or ATCC Deposit No. 98809; or (c) a nucleotide sequence complementary to any of the nucleotide sequences in (a) or (b). Additionally, an isolated nucleic acid of the invention may be a polynucleotide at least 90% but not more than 99% identical to (a) a nucleotide sequence encoding a *de novo* DNA cytosine methyltransferase polypeptide having the amino acid sequence in SEQ ID NO:4; (b) a nucleotide sequence encoding a *de novo* DNA cytosine methyltransferase polypeptide having the amino acid sequence encoded by the cDNA clone contained in ATCC Deposit No.326637; or (c) a nucleotide sequence complementary to any of the nucleotide sequences in (a) or (b).

Conventional means utilizing known computer programs such as the BestFit program (Wisconsin Sequence Analysis Package, Version 10 for Unix, Genetics Computer Group, University Research Park, 575 Science Drive,

Madison, WI 53711) may be utilized to determine if a particular nucleic acid molecule is at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% identical to any one of the nucleotide sequences shown in SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, or SEQ ID NO:4 or to any one of the nucleotide sequences of the deposited cDNA clones contained in ATCC Deposit No. 209933, ATCC Deposit No. 209934, ATCC Deposit No. 98809, or ATCC Deposit No. 326637.

Further preferred embodiments are polynucleotides encoding *de novo* DNA cytosine methyltransferases and *de novo* DNA cytosine methyltransferase variants that have an amino acid sequence of the *de novo* DNA cytosine methyltransferase protein of SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, or SEQ ID NO:8 in which several, 1, 1-2, 1-3, 1-5 or 5-10 amino acid residues are substituted, deleted or added, in any combination.

Further preferred embodiments of the invention are polynucleotides that are at least 90% identical over their entire length to a polynucleotide encoding a *de novo* DNA cytosine methyltransferase polypeptide having the amino acid sequence set out in SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, or SEQ ID NO:8, and polynucleotides which are complementary to such polynucleotides. Most highly preferred are polynucleotides that comprise regions that are at least 90% identical over their entire length to a polynucleotide encoding the *de novo* DNA cytosine methyltransferase polypeptides of the ATCC deposited human DNMT3A cDNA clone and polynucleotides complementary thereto, and 90% to 99% identical over their entire length to a polynucleotide encoding the *de novo* DNA cytosine methyltransferase polypeptides of the ATCC deposited human DNMT3B cDNA clone and polynucleotides complementary thereto. In this regard, polynucleotides at least 95% identical over their entire length to the same are particularly preferred, and those with at least 97% identity are especially preferred. Furthermore, those with at least 98% identity are highly preferred and with at least 99% identity being the most preferred.

In a more specific embodiment, the nucleic acid molecules of the present invention, *e.g.*, isolated nucleic acids comprising a polynucleotide having a

nucleotide sequence encoding a *de novo* DNA cytosine methyltransferase polypeptide or fragment thereof, are not the sequence of nucleotides, the nucleic acid molecules (*e.g.*, clones), or the nucleic acid inserts identified in one or more of the below cited public EST or STS GenBank Accession Reports.

The following public ESTs were identified that relate to portions of SEQ ID NO:1: AA052791(SEQ ID NO:9); AA111043(SEQ ID NO:10); AA154890(SEQ ID NO:11); AA240794(SEQ ID NO:12); AA756653(SEQ ID NO:13); W58898(SEQ ID NO:14); W59299(SEQ ID NO:15); W91664(SEQ ID NO:16); W91665(SEQ ID NO:17); to portions of SEQ ID NO:2: AA116694 (SEQ ID NO:18); AA119979 (SEQ ID NO:19); AA177277 (SEQ ID NO:20); AA210568 (SEQ ID NO:21); AA399749 (SEQ ID NO:22); AA407106 (SEQ ID NO:23); AA575617 (SEQ ID NO:24); to portions of SEQ ID NO:3: AA004310 (SEQ ID NO:25); AA004399 (SEQ ID NO:26); AA312013 (SEQ ID NO:27); AA355824 (SEQ ID NO:28); AA533619 (SEQ ID NO:29); AA361360 (SEQ ID NO:30); AA364876 (SEQ ID NO:31); AA503090 (SEQ ID NO:32); AA533619 (SEQ ID NO:33); AA706672 (SEQ ID NO:34); AA774277 (SEQ ID NO:35); AA780277 (SEQ ID NO:36); H03349 (SEQ ID NO:37); H04031 (SEQ ID NO:38); H53133 (SEQ ID NO:39); H53239 (SEQ ID NO:40); H64669 (SEQ ID NO:41); N26002 (SEQ ID NO:42); N52936 (SEQ ID NO:43); N88352 (SEQ ID NO:44); N89594 (SEQ ID NO:45); R19795 (SEQ ID NO:46); R47511 (SEQ ID NO:47); T50235 (SEQ ID NO:48); T78023 (SEQ ID NO:49); T78186 (SEQ ID NO:50); W22886 (SEQ ID NO:51); W67657 (SEQ ID NO:52); W68094 (SEQ ID NO:53); W76111 (SEQ ID NO:54); Z38299 (SEQ ID NO:55); Z42012 (SEQ ID NO:56); and that relate to SEQ ID NO:4: AA206103(SEQ ID NO:57); AA206264(SEQ ID NO:58); AA216527(SEQ ID NO:59); AA216697(SEQ ID NO:60); AA305044(SEQ ID NO:61); AA477705(SEQ ID NO:62); AA477706(SEQ ID NO:63); AA565566(SEQ ID NO:64); AA599893(SEQ ID NO:65); AA729418(SEQ ID NO:66); AA887508(SEQ ID NO:67); F09856(SEQ ID NO:68); F12227(SEQ ID NO:69); N39452(SEQ ID NO:70); N48564(SEQ ID NO:71); T66304(SEQ ID NO:72); and T66356(SEQ ID NO:73); AA736582(SEQ

ID NO:77); AA748883(SEQ ID NO:78); AA923295(SEQ ID NO:79); AAI000396(SEQ ID NO:80); AI332472(SEQ ID NO:81); W22473(SEQ ID NO:82) and the I.M.A.G.E. Consortium clone ID 22089 (ATCC Deposit No. 326637)(SEQ ID NO:76). Additionally, STSs G06200(SEQ ID NO:74) and G15302(SEQ ID NO:75) were identified in a search with SEQ ID NOS.:3 and 4, respectively.

The present invention is further directed to fragments of SEQ ID NO:1, 2 or 3, or to fragments of the cDNA nucleotide sequence found in ATCC Deposit Nos. 209933, 209934, or 98809. A fragment may be defined to be at least about 15 nt, and more preferably at least about 20 nt, still more preferably at least about 30 nt, and even more preferably, at least about 40 nt in length. Such fragments are useful as diagnostic probes and primers as discussed herein. Of course larger DNA fragments are also useful according to the present invention, as are fragments corresponding to most, if not all, of the nucleotide sequence of the cDNA clones contained in the plasmids deposited as ATCC Deposit No. 209933, ATCC Deposit No. 209934 or ATCC Deposit No. 98809;or as shown in SEQ ID NO:1, SEQ ID NO:2, or SEQ ID NO:3. Generally, polynucleotide fragments of the invention may be defined algebraically in the following way: (a) for SEQ ID NO:1, as 15 + N, wherein N equals zero or any positive integer up to 4176; (b) for SEQ ID NO:2, as 15 + N, wherein N equals zero or any positive integer up to 4180; and (c) for SEQ ID NO:3, as 15 + N, wherein N equals zero or any positive integer up to 4401. By a fragment at least 20 nt in length, for example, is intended fragments which include 20 or more contiguous bases from a nucleotide sequence of the ATCC deposited cDNAs or the nucleotide sequence as shown in SEQ ID NO:1, SEQ ID NO:2 or SEQ ID NO:3.

In another embodiment, the invention is directed to fragments of SEQ ID NO:4. Such fragments are defined as comprising the nucleotide sequence encoding the specific amino acid residues integral and immediately adjacent to the site where DNMT3B exons are spliced together. The DNMT3B sequence of SEQ ID NO:4 consists of 23 exon sequences defined accordingly: Exon 1 consists

of nucleotides 1-108 of SEQ ID NO:4; Exon 2 consists of nucleotides 109-256 of SEQ ID NO:4; Exon 3 consists of nucleotides 257-318 of SEQ ID NO:4; Exon 4 consists of nucleotides 319-420 of SEQ ID NO:4; Exon 5 consists of nucleotides 421-546 of SEQ ID NO:4; Exon 6 consists of nucleotides 547-768 of SEQ ID NO:4; Exon 7 consists of nucleotides 769-927 of SEQ ID NO:4; Exon 8 consists of nucleotides 928-1035 of SEQ ID NO:4; Exon 9 consists of nucleotides 1036-1180 of SEQ ID NO:4; Exon 10 consists of nucleotides 1181-1240 of SEQ ID NO:4; Exon 11 consists of nucleotides 1241-1366 of SEQ ID NO:4; Exon 12 consists of nucleotides 1367-1411 of SEQ ID NO:4; Exon 13 consists of nucleotide 1412-1491 of SEQ ID NO:4; Exon 14 consists of nucleotides 1492-1604 of SEQ ID NO:4; Exon 15 consists of nucleotides 1605-1788 of SEQ ID NO:4; Exon 16 consists of nucleotides 1789-1873 of SEQ ID NO:4; Exon 17 consists of nucleotides 1874-2019 of SEQ ID NO:4; Exon 18 consists of nucleotides 2020-2110 of SEQ ID NO:4; Exon 19 consists of nucleotides 2111-2259 of SEQ ID NO:4; Exon 20 consists of nucleotides 2260-2345 of SEQ ID NO:4; Exon 21 consists of nucleotides 2346-2415 of SEQ ID NO:4; Exon 22 consists of nucleotides 2416-2534 of SEQ ID NO:4; and Exon 23 consists of nucleotides 2535-4145 of SEQ ID NO:4.

It should be understood by those skilled in the art that with regards to SEQ ID NO:4, Exon 1 and Exon 23 are herein defined for the purposes of the invention. The first nucleotide of Exon 1 may or may not be the transcriptional start site for the DNMT3B genomic locus, and the last nucleotide identified for Exon 23 may or may not reflect the last nucleotide transcribed *in vivo*.

Thus, by way of example, fragments of SEQ ID NO:4 comprise the following exon-exon junctions of 20 nucleotides in length: the exon1/exon 2 junction of nucleotides 98-118 of SEQ ID NO:4; the exon 2/exon 3 junction of nucleotides 246-266 of SEQ ID NO:4; the exon 3/exon 4 junction of nucleotides 308-328 of SEQ ID NO:4; the exon 4/exon 5 junction of nucleotides 410-430 of SEQ ID NO:4; the exon 5/exon 6 junction of nucleotides 536-556 of SEQ ID NO:4; the exon 6/exon 7 junction of nucleotides 758-778 of SEQ ID NO:4; the

exon 7/exon 8 junction of nucleotides 917-937 of SEQ ID NO:4; the exon 8/exon 9 junction of nucleotides 1025-1045 of SEQ ID NO:4; the exon 9/exon 10 junction of nucleotides 1170-1190 of SEQ ID NO:4; the exon 10/exon 11 junction of nucleotides 1230-1250 of SEQ ID NO:4; the exon 11/exon 12 junction of nucleotides 1356-1376 of SEQ ID NO:4; the exon 12/exon 13 junction of nucleotides 1401-1421 of SEQ ID NO:4; the exon 13/exon 14 junction of nucleotides 1481-1501 of SEQ ID NO:4; the exon 14/exon 15 junction of nucleotides 1594-1614 of SEQ ID NO:4; the exon 15/exon 16 junction of nucleotides 1778-1798 of SEQ ID NO:4; the exon 16/exon 17 junction of nucleotides 1863-1883 of SEQ ID NO:4; the exon 17/exon 18 junction of nucleotides 2009-2029 of SEQ ID NO:4; the exon 18/exon 19 junction of nucleotides 2100-2120 of SEQ ID NO:4; the exon 19/exon 20 junction of nucleotides 2249-2269 of SEQ ID NO:4; the exon 20/exon 21 junction of nucleotides 2335-2355 of SEQ ID NO:4; the exon 21/exon 22 junction of nucleotides 2405-2425 of SEQ ID NO:4; and the exon 22/exon 23 junction of nucleotides 2524-2544 of SEQ ID NO:4.

As will be clear to those skilled in the art, other exon-exon junction fragments of SEQ ID NO:4 are possible which comprise 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, etc., nucleotides of SEQ ID NO:4. For the purposes of constructing such fragments, the following exon-exon junctions are identified: the exon1/exon 2 junction of nucleotides 108 and 109 of SEQ ID NO:4; the exon 2/exon 3 junction of nucleotides 256 and 257 of SEQ ID NO:4; the exon 3/exon 4 junction of nucleotides 318 and 319 of SEQ ID NO:4; the exon 4/exon 5 junction of nucleotides 420 and 421 of SEQ ID NO:4; the exon 5/exon 6 junction of nucleotides 546 and 547 of SEQ ID NO:4; the exon 6/exon 7 junction of nucleotides 768 and 769 of SEQ ID NO:4; the exon 7/exon 8 junction of nucleotides 927 and 928 of SEQ ID NO:4; the exon 8/exon 9 junction of nucleotides 1035 and 1036 of SEQ ID NO:4; the exon 9/exon 10 junction of nucleotides 1180 and 1181 of SEQ ID NO:4; the exon 10/exon 11 junction of nucleotides 1240 and 1241 of SEQ ID NO:4; the exon 11/exon 12 junction of

nucleotides 1366 and 1367 of SEQ ID NO:4; the exon 12/exon 13 junction of nucleotides 1411 and 1412 of SEQ ID NO:4; the exon 13/exon 14 junction of nucleotides 1491 and 1492 of SEQ ID NO:4; the exon 14/exon 15 junction of nucleotides 1604 and 1605 of SEQ ID NO:4; the exon 15/exon 16 junction of nucleotides 1788 and 1789 of SEQ ID NO:4; the exon 16/exon 17 junction of nucleotides 1873 and 1874 of SEQ ID NO:4; the exon 17/exon 18 junction of nucleotides 2019 and 2020 of SEQ ID NO:4; the exon 18/exon 19 junction of nucleotides 2110 and 2111 of SEQ ID NO:4; the exon 19/exon 20 junction of nucleotides 2259 and 2260 of SEQ ID NO:4; the exon 20/exon 21 junction of nucleotides 2345 and 2346 of SEQ ID NO:4; the exon 21/exon 22 junction of nucleotides 2415 and 2416 of SEQ ID NO:4; and the exon 22/exon 23 junction of nucleotides 2534 and 2535 of SEQ ID NO:4. Junction nucleotides may be located at any position of the selected SEQ ID NO:4 fragment.

The present invention further relates to polynucleotides that hybridize to the above-described sequences. In this regard, the present invention especially relates to polynucleotides that hybridize under stringent conditions to the above-described polynucleotides. As herein used, the term "stringent conditions" means hybridization will occur only if there is at least 90% and preferably at least 95% identity and more preferably at least 97% identity between the sequences.

Furthermore, a major consideration associated with hybridization analysis of DNA or RNA sequences is the degree of relatedness the probe has with the sequences present in the specimen under study. This is important with a blotting technique (e.g., Southern or Northern Blot), since a moderate degree of sequence homology under nonstringent conditions of hybridization can yield a strong signal even though the probe and sequences in the sample represent non-homologous genes.

The particular hybridization technique is not essential to the invention, any technique commonly used in the art is within the scope of the present invention. Typical probe technology is described in United States Patent 4,358,535 to Falkow *et al.*, incorporated by reference herein. For example,

hybridization can be carried out in a solution containing 6 x SSC (10 x SSC: 1.5 M sodium chloride, 0.15 M sodium citrate, pH 7.0), 5 x Denhardt's (1 x Denhardt's: 0.2% bovine serum albumin, 0.2% polyvinylpyrrolidone, 0.02% Ficoll 400), 10 mM EDTA, 0.5% SDS and about 10^7 cpm of nick-translated DNA for 16 hours at 65°C. Additionally, if hybridization is to an immobilized nucleic acid, a washing step may be utilized wherein probe binding to polynucleotides of low homology, or nonspecific binding of the probe, may be removed. For example, a stringent wash step may involve a buffer of 0.2 x SSC and 0.5% SDS at a temperature of 65°C.

Additional information related to hybridization technology and, more particularly, the stringency of hybridization and washing conditions may be found in Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*, Second Edition, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York (1989), which is incorporated herein by reference.

Polynucleotides of the invention which are sufficiently identical to a nucleotide sequences contained in SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3 or SEQ ID NO:4, or in the cDNA inserts of ATCC Deposit No. 209933, ATCC Deposit No. 209934, ATCC Deposit No. 98809 or ATCC Deposit No. 326637, may be used as hybridization probes for cDNA and genomic DNA, to isolate full-length cDNAs and genomic clones encoding *de novo* DNA cytosine methyltransferase proteins and to isolate cDNA and genomic clones of other genes that have a high sequence similarity to the *de novo* DNA cytosine methyltransferase genes. Such hybridization techniques are known to those of skill in the art. Typically, these nucleotide sequences are at least about 90% identical, preferably at least about 95% identical, more preferably at least about 97%, 98% or 99% identical to that of the reference. The probes generally will comprise at least 15 nucleotides. Preferably, such probes will have at least 30 nucleotides and may have at least 50 nucleotides. Particularly preferred probes will range between 30 and 50 nucleotides.

The polynucleotides and polypeptides of the present invention may be employed as research reagents and materials for discovery of treatments and diagnostics to animal and human disease.

III. Vectors, Host Cells, and Recombinant Expression

The present invention also relates to vectors that comprise a polynucleotide of the present invention, host cells which are genetically engineered with vectors of the invention and the production of polypeptides of the invention by recombinant techniques. Cell-free translation systems can also be employed to produce such proteins using RNAs derived from the DNA constructs of the invention.

For recombinant production, host cells can be genetically engineered to incorporate expression systems for polynucleotides of the invention. Introduction of polynucleotides into host cells can be effected by methods described in many standard laboratory manuals, such as Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*, 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989). For example, calcium phosphate transfection, DEAE-dextran mediated transfection, transvection, microinjection, cationic lipid-mediated transfection, electroporation, transduction, scrape loading, ballistic introduction, infection or any other means known in the art may be utilized.

Representative examples of appropriate hosts include bacterial cells, such as streptococci, staphylococci, *E. coli*, *Streptomyces* and *Bacillus subtilis* cells; fungal cells, such as yeast cells and *Aspergillus* cells; insect cells such as *Drosophila* S2 and *Spodoptera* Sf9 cells; animal cells such as CHO, COS, HeLa, C127, 3T3, BHK, 293 and Bowes melanoma cells; and plant cells.

A great variety of expression systems can be used. Such systems include, among others, chromosomal, episomal and virus-derived systems, *e.g.*, vectors derived from bacterial plasmids, from bacteriophages, from transposons, from yeast episomes, from insertion elements, from yeast chromosomal elements, from

viruses such as baculoviruses, papova viruses, such as SV40, vaccinia viruses, adenoviruses, fowl pox viruses, pseudorabies viruses, and retroviruses, and vectors derived from combinations thereof, such as those derived from plasmid and bacteriophage genetic elements, such as cosmids and phagemids. The expression systems may contain control regions that regulate as well as engender expression. Generally, any system or vector suitable to maintain, propagate or express polynucleotides to produce a polypeptide in a host may be used. The appropriate nucleotide sequence may be inserted into an expression system by any of a variety of well-known and routine techniques, such as, for example, those set forth in Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual (supra)*.

RNA vectors may also be utilized for the expression of the *de novo* DNA cytosine methyltransferases disclosed in this invention. These vectors are based on positive or negative strand RNA viruses that naturally replicate in a wide variety of eukaryotic cells (Bredenbeek, P.J. and Rice, C.M., Virology 3: 297-310, (1992)). Unlike retroviruses, these viruses lack an intermediate DNA life-cycle phase, existing entirely in RNA form. For example, alpha viruses are used as expression vectors for foreign proteins because they can be utilized in a broad range of host cells and provide a high level of expression; examples of viruses of this type include the Sindbis virus and Semliki Forest virus (Schlesinger, S., TIBTECH 11: 18-22, (1993); Frolov, I., *et al.*, *Proc. Natl. Acad. Sci. (USA)* 93: 11371-11377, (1996)). As exemplified by Invitrogen's Sinbis expression system, the investigator may conveniently maintain the recombinant molecule in DNA form (pSinrep5 plasmid) in the laboratory, but propagation in RNA form is feasible as well. In the host cell used for expression, the vector containing the gene of interest exists completely in RNA form and may be continuously propagated in that state if desired.

For secretion of the translated protein into the lumen of the endoplasmic reticulum, into the periplasmic space or into the extracellular environment appropriate secretion signals may be incorporated into the desired polypeptide.

These signals may be endogenous to the polypeptide or they may be heterologous signals.

As used herein, the term "operably linked," when used in the context of a linkage between a structural gene and an expression control sequence, *e.g.*, a promoter, refers to the position and orientation of the expression control sequence relative to the structural gene so as to permit expression of the structural gene in any host cell. For example, an operable linkage would maintain proper reading frame and would not introduce any in frame stop codons.

As used herein, the term "heterologous promoter," refers to a promoter not normally and naturally associated with the structural gene to be expressed. For example, in the context of expression of a *de novo* DNA cytosine methyltransferase polypeptide, a heterologous promoter would be any promoter other than an endogenous promoter associated with the *de novo* DNA cytosine methyltransferase gene in non-recombinant mouse or human chromosomes. In specific embodiments of this invention, the heterologous promoter is a prokaryotic or bacteriophage promoter, such as the lac promoter, T3 promoter, or T7 promoter. In other embodiments, the heterologous promoter is a eukaryotic promoter.

In other embodiments, this invention provides an isolated nucleic acid molecule comprising a *de novo* DNA cytosine methyltransferase structural gene operably linked to a heterologous promoter. As used herein, the term "a *de novo* DNA cytosine methyltransferase structural gene" refers to a nucleotide sequence at least about 90% identical to one of the following nucleotide sequences:

(a) a nucleotide sequence encoding the *de novo* DNA cytosine methyltransferase polypeptide having the complete amino acid sequence in SEQ ID NO:5, SEQ ID NO:6, or SEQ ID NO:7;

(b) a nucleotide sequence encoding the *de novo* DNA cytosine methyltransferase polypeptide having the complete amino acid sequence encoded by the cDNA insert of ATCC Deposit No. 209933, ATCC Deposit No. 209934, or ATCC Deposit No. 98809; or

(c) a nucleotide sequence complementary to any of the nucleotide sequences in (a) or (b).

In preferred embodiments, the *de novo* DNA cytosine methyltransferase structural gene is 90%, and more preferably 91%, 92%, 93%, 94%, 95%, 97%, 98%, 99%, or 100% identical to one or more of nucleotide sequences (a), (b), or (c) *supra*.

In another embodiment the term "a *de novo* DNA cytosine methyltransferase structural gene" refers to a nucleotide sequence about 90% to 99% identical to one of the following nucleotide sequences:

(a) a nucleotide sequence encoding the *de novo* DNA cytosine methyltransferase polypeptide having the complete amino acid sequence in SEQ ID NO:8;

(b) a nucleotide sequence encoding the *de novo* DNA cytosine methyltransferase polypeptide having the complete amino acid sequence encoded by the cDNA insert of ATCC Deposit No. 326637; or

(c) a nucleotide sequence complementary to any of the nucleotide sequences in (a) or (b).

In preferred embodiments, the *de novo* DNA cytosine methyltransferase structural gene is 90%, and more preferably 91%, 92%, 93%, 94%, 95%, 97%, 98%, or 99% identical to SEQ ID NO:8, ATCC Deposit No. 326637 or polynucleotides complementary thereto.

This invention also provides an isolated nucleic acid molecule comprising a *de novo* DNA cytosine methyltransferase structural gene operably linked to a heterologous promoter, wherein said isolated nucleic acid molecule does not encode a fusion protein comprising the *de novo* DNA cytosine methyltransferase structural gene or a fragment thereof.

This invention further provides an isolated nucleic acid molecule comprising a *de novo* DNA cytosine methyltransferase structural gene operably linked to a heterologous promoter, wherein said isolated nucleic acid molecule

is capable of expressing a *de novo* DNA cytosine methyltransferase polypeptide when used to transform an appropriate host cell.

This invention also provides an isolated nucleic acid molecule comprising a polynucleotide having a nucleotide sequence at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to a sequence encoding a *de novo* DNA cytosine methyltransferase polypeptide having the amino acid sequence of SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7 or SEQ ID NO:8, wherein said isolated nucleic acid molecule does not contain a nucleotide sequence at least 90% identical to the 3' untranslated region of SEQ ID NO:1 (nucleotides 2942-4191), SEQ ID NO:2 (nucleotides 2847-4174), SEQ ID NO:3 (nucleotides 3090-4397) or SEQ ID NO:4 (nucleotides 2677-4127), or a fragment of the 3' untranslated region greater than 25, 50, 75, 100, or 125 bp in length.

This invention further provides an isolated nucleic acid molecule comprising a polynucleotide having a nucleotide sequence at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to a sequence encoding a *de novo* DNA cytosine methyltransferase polypeptide having the amino acid sequence of SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7 or SEQ ID NO:8, wherein said isolated nucleic acid molecule does not contain a nucleotide sequence at least 90% identical to the 5' untranslated region of SEQ ID NO:1 (nucleotides 1-216), SEQ ID NO:2 (nucleotides 1-268), SEQ ID NO:3 (nucleotides 1-352) or SEQ ID NO:4 (nucleotides 1-114), or a fragment of the 5' untranslated region greater than 25, 35, 45, 55, 65, 75, 85, or 90 bp.

Suitable known prokaryotic promoters for use in the production of proteins of the present invention include the *E. coli lac*I and *lac*Z promoters, the T3 and T7 promoters, the *gpt* promoter, the lambda PR and PL promoters and the *trp* promoter. Suitable eukaryotic promoters include the CMV immediate early promoter, the HSV thymidine kinase promoter, the early and late SV40 promoters, the promoters of retroviral LTRs, such as those of the Rous Sarcoma Virus (RSV), adenovirus promoter, *Herpes* virus promoter, and metallothionein

promoters, such as the mouse metallothionein-I promoter and tissue and organ-specific promoters known in the art.

If the *de novo* DNA cytosine methyltransferase polypeptide is to be expressed for use in screening assays, generally, it is preferred that the polypeptide be produced at the surface of the cell. In this event, the cells may be harvested prior to use in the screening assay. If *de novo* DNA cytosine methyltransferase polypeptide is secreted into the medium, the medium can be recovered in order to recover and purify the polypeptide; if produced intracellularly, the cells must first be lysed before the polypeptide is recovered.

De novo DNA cytosine methyltransferase polypeptides can be recovered and purified from recombinant cell cultures by well-known methods including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography and lectin chromatography. Most preferably, high performance liquid chromatography is employed for purification. Well known techniques for refolding proteins may be employed to regenerate active conformation when the polypeptide is denatured during isolation and or purification.

IV. Polypeptides of the Invention

The *de novo* DNA cytosine methyltransferase polypeptides of the present invention include the polypeptide of SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7 or SEQ ID NO:8, as well as polypeptides and fragments which have activity and have at least 90% identity to the polypeptide of SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7 or SEQ ID NO:8, or the relevant portion and more preferably at least 96%, 97% or 98% identity to the polypeptide of SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7 or SEQ ID NO:8, and still more preferably at least 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identity to the polypeptide of SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7 or SEQ ID NO:8.

The polypeptides of the present invention are preferably provided in an isolated form.

The polypeptides of the present invention include the polypeptide encoded by the deposited cDNAs; a polypeptide comprising amino acids from about 1 to about 908 in SEQ ID NO:5; a polypeptide comprising amino acids from about 1 to about 859 in SEQ ID NO:6; a polypeptide comprising amino acids from about 1 to about 912 in SEQ ID NO:7 and a polypeptide comprising amino acids from about 1 to about 853 in SEQ ID NO:8; as well as polypeptides which are at least about 90% identical, and more preferably at least about 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to the polypeptides described above and also include portions of such polypeptides with at least 30 amino acids and more preferably at least 50 amino acids.

Polypeptides of the invention also include alternative splicing variants of the Dnmt3 sequences disclosed herein. For example, alternative variant spliced proteins of mouse Dnmt3b include but are not limited to a polypeptide wherein, except for at least one conservative amino acid substitution, said polypeptide has a sequence selected from the group consisting of: (1) amino acid residues 1 to 362 and 383 to 859 from SEQ ID NO:2; and (2) amino acid residues 1 to 362 and 383 to 749 and 813 to 859 from SEQ ID NO:2; and alternative variant spliced proteins of human DNMT3B include but are not limited to a polypeptide wherein, except for at least one conservative amino acid substitution, said polypeptide has a sequence selected from the group consisting of: (1) amino acid residues 1 to 355 and 376 to 853 from SEQ ID NO:4; and (2) amino acid residues 1 to 355 and 376 to 743 and 807 to 853 from SEQ ID NO:4.

The *de novo* DNA cytosine methyltransferase polypeptides may be a part of a larger protein such as a fusion protein. It is often advantageous to include additional amino acid sequence which contains secretory or leader sequences, pro-sequences, sequences which aid in purification such as multiple histidine residues, or additional sequence for stability during recombinant production.

Biologically active fragments of the *de novo* DNA cytosine methyltransferase polypeptides are also included in the invention. A fragment is a polypeptide having an amino acid sequence that entirely is the same as part but not all of the amino acid sequence of one of the aforementioned *de novo* DNA cytosine methyltransferase polypeptides. As with *de novo* DNA cytosine methyltransferase polypeptides, fragments may be "free-standing," or comprised within a larger polypeptide of which they form a part or region, most preferably as a single continuous region. In the context of this invention, a fragment may constitute from about 10 contiguous amino acids identified in SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7 or SEQ ID NO:8. More specifically, polypeptide fragment lengths may be defined algebraically as follows: (a) for SEQ ID NO:5, as 10 + N, wherein N equals zero or any positive integer up to 898; (b) for SEQ ID NO:6, as 10 + N, wherein N equals zero or any positive integer up to 849; (c) for SEQ ID NO:7, as 10 + N, wherein N equals zero or any positive integer up to 902; and (d) for SEQ ID NO:8, as 10 + N, wherein N equals zero or any positive integer up to 843.

Preferred fragments include, for example, truncation polypeptides having the amino acid sequence of *de novo* DNA cytosine methyltransferase polypeptides, except for deletion of a continuous series of residues that includes the amino terminus, or a continuous series of residues that includes the carboxyl terminus or deletion of two continuous series of residues, one including the amino terminus and one including the carboxyl terminus. Also preferred are fragments characterized by structural or functional attributes such as fragments that comprise alpha-helix and alpha-helix forming regions, beta-sheet and beta-sheet-forming regions, turn and turn-forming regions, coil and coil-forming regions, hydrophilic regions, hydrophobic regions, alpha amphipathic regions, beta amphipathic regions, flexible regions, surface-forming regions, substrate binding region, and high antigenic index regions. Biologically active fragments are those that mediate protein activity, including those with a similar activity or an

improved activity, or with a decreased undesirable activity. Also included are those that are antigenic or immunogenic in an animal, especially in a human.

Thus, the polypeptides of the invention include polypeptides having an amino acid sequence at least 90% identical to that of SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7 or SEQ ID NO:8, or fragments thereof with at least 90% identity to the corresponding fragment of SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7 or SEQ ID NO:8, all of which retain the biological activity of the *de novo* DNA cytosine methyltransferase protein, including antigenic activity. Included in this group are variants of the defined sequence and fragment. Preferred variants are those that vary from the reference by conservative amino acid substitutions, *i.e.*, those that substitute a residue with another of like characteristics. Typical substitutions are among Ala, Val, Leu and Ile; among Ser and Thr; among the acidic residues Asp and Glu; among Asn and Gln; and among the basic residues Lys and Arg, or aromatic residues Phe and Tyr. Particularly preferred are variants in which several, 5 to 10, 1 to 5, or 1 to 2 amino acids are substituted, deleted, or added in any combination.

The *de novo* DNA cytosine methyltransferase polypeptides of the invention can be prepared in any suitable manner. Such polypeptides include isolated naturally occurring polypeptides, recombinantly produced polypeptides, synthetically produced polypeptides, or polypeptides produced by a combination of these methods. Means for preparing such polypeptides are well understood in the art.

V. In Vitro DNA Methylation

One preferred embodiment of the invention enables the *in vitro* methylation at the C5 position of cytosine in DNA. The starting substrate DNA may be hemimethylated (i.e., one strand of the duplex DNA is methylated) or may lack methylation completely. The polypeptides of the invention, being *de novo* DNA cytosine methyltransferases, are uniquely suited to the latter function, owing

to the fact that, unlike maintenance methyltransferases, their preferred substrate is not hemimethylated DNA.

As exemplified in Examples 7 and 8, isolated polypeptides of the invention function as *in vitro* DNA methyltransferases when combined in an appropriately buffered solution with the appropriate cofactors and a substrate DNA. The substrate DNA may be selected from any natural source, e.g., genomic DNA, or a recombinant source such as a DNA fragment amplified by the polymerase chain reaction. The substrate DNA may be prokaryotic or eukaryotic DNA. In a preferred embodiment, the substrate DNA is mammalian DNA, and most preferredly, the substrate DNA is human DNA.

It will be well appreciated by those in the art that *in vitro* methylation of DNA may be used to direct or regulate the expression of said DNA in a biological system. For example, over-expression, under-expression or lack of expression of a particular native DNA sequence in a host cell or organism may be attributed to the fact that the DNA is under-methylated (hypomethylated) or not methylated. Thus, *in vitro* methylation of a recombinant form of said DNA, and the subsequent introduction of the methylated, recombinant DNA into the cell or organism, may effect an increase or decrease in the expression of the encoded polypeptide.

Also, it will be readily apparent to the skilled artisan that the *in vitro* methylation pattern will be maintained after introduction into a biological system by the action of maintenance methyltransferase polypeptides in said system.

In one embodiment of the invention, the biological system selected for the introduction of *in vitro* methylated DNA may be prokaryotic or eukaryotic. In a preferred embodiment, the biological system is mammalian, and the most preferred embodiment is when the biological system is human.

Methods for introducing the *in vitro* methylated DNA into the biological system are well known in the art, and the skilled artisan will recognize that the *in vitro* methylation of DNA may be a preliminary step to any system of gene therapy detailed herein.

VI. Genetic Screening and Diagnostic Assays

To map the human chromosome locations, the GenBank STS database was searched using Dnmt3a and Dnmt3b sequences as queries. The search identified markers WI-6283 (GenBank Accession number G06200) and SHGC-15969 (GenBank Accession number G15302) as matching the cDNA sequence of Dnmt3a and Dnmt3b, respectively. WI-6283 has been mapped to 2p23 between D2S171 and D2S174 (48-50 cM) on the radiation hybrid map by Whitehead Institute/MIT Center for Genome Research. The corresponding mouse chromosome location is at 4.0 cM on chromosome 12. SHGC-15969 has been mapped to 20pl 1.2 between D20S184 and D20S106 (48-50 cM) by Stanford Human Genome Center. The corresponding mouse chromosome locus is at 84.0 cM on chromosome 2.

These data are valuable as markers to be correlated with genetic map data. Such data are found, for example, in V. McKusick, Mendelian Inheritance in Man (available on-line through Johns Hopkins, University Welch Medical Library). The relationship between genes and diseases that have been mapped to the same chromosomal region are then identified through linkage analysis (coinheritence of physically adjacent genes).

The differences in the cDNA or genomic sequence between affected and unaffected individuals can also be determined. If a mutation is observed in some or all of the affected individuals but not in any normal individuals, then the mutation is likely to be the causative agent of the disease.

This invention also relates to the use of *de novo* DNA cytosine methyltransferase polynucleotides for use as diagnostic reagents. Detection of a mutated form of a *de novo* DNA cytosine methyltransferase gene associated with a dysfunction will provide a diagnostic tool that can add to or define a diagnosis of a disease or susceptibility to a disease which results from under-expression, over-expression or altered expression of the mutated *de novo* DNA cytosine methyltransferase. Individuals carrying mutations in one or more *de novo* DNA

cytosine methyltransferase genes may be detected at the DNA level by a variety of techniques.

Nucleic acids for diagnosis may be obtained from a subject's cells, such as from blood, urine, saliva, tissue biopsy or autopsy material. The genomic DNA may be used directly for detection or may be amplified enzymatically by using PCR or other amplification techniques prior to analysis. RNA or cDNA may also be used in similar fashion. Deletions and insertions can be detected by a change in size of the amplified product in comparison to the normal genotype. Point mutations can be identified by hybridizing amplified DNA to labeled *de novo* DNA cytosine methyltransferase nucleotide sequences. Perfectly matched sequences can be distinguished from mismatched duplexes by RNase digestion or by differences in melting temperatures. DNA sequence differences may also be detected by alterations in electrophoretic mobility of DNA fragments in gels, with or without denaturing agents, or by direct DNA sequencing (see, e.g., Myers, *et al.*, *Science 230*:1242 (1985)). Sequence changes at specific locations may also be revealed by nuclease protection assays, such as RNase and S1 protection or the chemical cleavage method (see Cotton, *et al.*, *Proc. Natl. Acad. Sci. USA 85*:4397-4401 (1985)).

The diagnostic assays offer a process for diagnosing or determining a susceptibility to neoplastic disorders through detection of mutations in one or more *de novo* DNA cytosine methyltransferase genes by the methods described.

In addition, neoplastic disorders may be diagnosed by methods that determine an abnormally decreased or increased level of *de novo* DNA cytosine methyltransferase polypeptide or *de novo* DNA cytosine methyltransferase mRNA in a sample derived from a subject. Decreased or increased expression may be measured at the RNA level using any of the methods well known in the art for the quantitation of polynucleotides; for example, RT-PCR, RNase protection, Northern blotting and other hybridization methods may be utilized. Assay techniques that may be used to determine the level of a protein, such as an *de novo* DNA cytosine methyltransferase protein, in a sample derived from a host

are well known to those of skill in the art. Such assay methods include radioimmunoassays, competitive-binding assays, Western blot analysis and ELISA assays.

Additionally, methods are provided for diagnosing or determining a susceptibility of an individual to neoplastic disorders, comprising (a) assaying the *de novo* DNA cytosine methyltransferase protein gene expression level in mammalian cells or body fluid; and (b) comparing said *de novo* DNA cytosine methyltransferase protein gene expression level with a standard *de novo* DNA cytosine methyltransferase protein gene expression level whereby an increase or decrease in said *de novo* DNA cytosine methyltransferase gene expression level over said standard is indicative of an increased or decreased susceptibility to a neoplastic disorder.

VII. De novo DNA Cytosine Methyltransferase Antibodies

The polypeptides of the invention or their fragments or analogs thereof, or cells expressing them may also be used as immunogens to produce antibodies immunospecific for the *de novo* DNA cytosine methyltransferase polypeptides. By "immunospecific" is meant that the antibodies have affinities for the polypeptides of the invention that are substantially greater in their affinities for related polypeptides such as the analogous proteins of the prior art.

Antibodies generated against the *de novo* DNA cytosine methyltransferase polypeptides can be obtained by administering the polypeptides or epitope-bearing fragments, analogs or cells to an animal, preferably a nonhuman, using routine protocols. For preparation of monoclonal antibodies, any technique which provides antibodies produced by continuous cell line cultures can be used. Examples include the hybridoma technique (Kohler, G. and Milstein, C., Nature 256:495-497 (1975)), the trioma technique, the human B-cell hybridoma technique (Kozbor, *et al.*, *Immunology Today 4*:72 (1983)) and the EBV-

hybridoma technique (Cole, *et al.*, *Monoclonal Antibodies and Cancer Therapy*, pp. 77-96, Alan R. Liss, Inc., (1985)).

Techniques for the production of single chain antibodies (U.S. Patent No. 4,946,778) may also be adapted to produce single chain antibodies to polypeptides of this invention. Also, transgenic mice, or other organisms including other mammals, may be used to express humanized antibodies.

The above-described antibodies may be employed to isolate or to identify clones expressing the polypeptide or to purify the polypeptides by affinity chromatography.

Antibodies against *de novo* DNA cytosine methyltransferase polypeptides may also be employed to treat neoplastic disorders, among others.

VIII. Agonist and Antagonist Screening

The *de novo* DNA cytosine methyltransferase polypeptides of the present invention may be employed in a screening process for compounds which bind one of the proteins and which activate (agonists) or inhibit activation of (antagonists) one of the polypeptides of the present invention. Thus, polypeptides of the invention may also be used to assess the binding of small molecule substrates and ligands in, for example, cells, cell-free preparations, chemical libraries, and natural product mixtures. These substrates and ligands may be natural substrates and ligands or may be structural or functional mimetics (see Coligan, *et al.*, *Current Protocols in Immunology 1*(2):Chapter 5 (1991)).

By "agonist" is intended naturally occurring and synthetic compounds capable of enhancing a *de novo* DNA cytosine methyltransferase activity (e.g., increasing the rate of DNA methylation). By "antagonist" is intended naturally occurring and synthetic compounds capable of inhibiting a *de novo* DNA cytosine methyltransferase activity.

DNA methylation is an important, fundamental regulatory mechanism for gene expression, and, therefore, the methylated state of a particular DNA

sequence may be associated with many pathologies. Accordingly, it is desirous to find both compounds and drugs which stimulate *de novo* DNA cytosine methyltransferase activity and which can inhibit the function of *de novo* DNA cytosine methyltransferase protein. In general, agonists are employed for therapeutic and prophylactic purposes including the treatment of ceratin types of neoplastic disorders. For example, *de novo* methylation of growth regulatory genes in somatic tissues is associated with tumorigenesis in humans (Laird, P. W. and Jaenisch, R. *Ann. Rev. Genet. 30*:441-464 (1996); Baylin, S. B. *et al., Adv. Cancer. Res. 72*:141-196 (1998); and Jones, P. A. and Gonzalgo, M. L. *Proc. Natl. Acad. Sci. USA 94*:2103-2105 (1997)).

In general, such screening procedures involve producing appropriate cells which express the polypeptide of the present invention. Such cells include cells from mammals, yeast, *Drosophila* or *E. coli.* Cells expressing the protein (or cell membrane containing the expressed protein) are then contacted with a test compound to observe binding, stimulation or inhibition of a functional response.

Alternatively, the screening procedure may be an *in vitro* procedure in which the activity of isolated DNMT3 protein is tested in the presence of a potential agonist or antagonist of DNMT3 *de novo* DNA cytosine methyltransferase activity. Such in vitro assays are known to those skilled in the art, and by way of example are demonstrated in Example 4.

The assays may simply test binding of a candidate compound wherein adherence to the cells bearing the protein is detected by means of a label directly or indirectly associated with the candidate compound or in an assay involving competition with a labeled competitor. Further, these assays may test whether the candidate compound affects activity of the protein, using detection systems appropriate to the cells bearing the protein at their surfaces. Inhibitors of activation are generally assayed in the presence of a known agonist and the effect on activation by the agonist in the presence of the candidate compound is observed. Standard methods for conducting such screening assays are well understood in the art.

Examples of potential *de novo* DNA cytosine methyltransferase protein antagonists include antibodies or, in some cases, oligonucleotides or proteins which are closely related to the substrate of the *de novo* DNA cytosine methyltransferase protein, e.g., small molecules which bind to the protein so that the activity of the protein is prevented.

IX. Gene Therapy Applications

For overview of gene therapy, see Strachan, T. & Read A.P., Chapter 20, "Gene Therapy and Other Molecular Genetic-based Therapeutic Approaches," (and references cited therein) in *Human Molecular Genetics*, BIOS Scientific Publishers Ltd. (1996).

Initial research in the area of gene therapy focused on a few well-characterized and highly publicized disorders: cystic fibrosis (Drumm, M.L. *et al.*, *Cell 62*:1227-1233 (1990); Gregory, R.J. *et al.*, *Nature 347*:358-363 (1990); Rich, D.P. *et al.*, *Nature 347*:358-363 (1990)); and Gaucher disease (Sorge, J. *et al.*, *Proc. Natl. Acad. Sci. (USA) 84*:906-909 (1987); Fink, J.K. *et al.*, *Proc. Natl. Acad. Sci. (USA) 87*:2334-2338 (1990)); and certain forms of hemophilia-Bontempo, F.A. *et al.*, *Blood 69*:1721-1724 (1987); Palmer, T.D. *et al.*, *Blood 73*:438-445 (1989); Axelrod, J.H. *et al.*, *Proc. Natl. Acad. Sci. (USA) 87*:5173-5177 (1990); Armentano, D. *et al.*, *Proc. Natl. Acad. Sci. (USA) 87*:6141-6145 (1990)); and muscular dystrophy (Partridge, T.A. *et al.*, *Nature 337*:176-179 (1989); Law, P.K. *et al.*, *Lancet 336*:114-115 (1990); Morgan, J.E. *et al.*, *J. Cell Biol. 111*:2437-2449 (1990)).

More recently, the application of gene therapy in the treatment of a wider variety of disorders is progressing, for example: cancer (Runnebaum, I.B., *Anticancer Res.* 17(4B): 2887-2890, (1997)), heart disease (Rader, D.J., *Int. J. Clin. Lab. Res.* 27(1): 35-43, (1997); Malosky, S., *Curr. Opin. Cardiol.* 11(4): 361-368, (1996)), central nervous system disorders and injuries (Yang, K., *et al.*, *Neurotrauma J.* 14(5): 281-297, (1997); Zlokovic, B.V., *et al.*, *Neurosurgery*

40(4): 789-803, (1997); Zlokovic, B.V., *et al.*, *Neurosurgery* 40(4): 805-812, (1997)), vascular diseases (Clowes, A.W., *Thromb. Haemost.* 78(1): 605-610, 1997), muscle disorders (Douglas, J.T., *et al.*, *Neuromuscul. Disord.* 7(5): 284-298, (1997); Huard, J., *et al.*, *Neuromuscul. Disord.* 7(5): 299-313, (1997)), rheumatoid arthritis (Evans, C.H., *et al.*, *Curr. Opin. Rheumatol.* 8(3): 230-234, (1996)) and epithelial tissue disorders (Greenhalgh, D.A., *et al.*, *Invest Dermatol. J.* 103(5 Suppl.): 63S-93S, (1994)).

In a preferred approach, one or more isolated nucleic acid molecules of the invention are introduced into or administered to the animal. Such isolated nucleic acid molecules may be incorporated into a vector or virion suitable for introducing the nucleic acid molecules into the cells or tissues of the animal to be treated, to form a transfection vector. Techniques for the formation of vectors or virions comprising the *de novo* DNA cytosine methyltransferase-encoding nucleic acid molecules are well known in the art and are generally described in "Working Toward Human Gene Therapy," Chapter 28 in *Recombinant DNA, 2nd Ed.*, Watson, J.D. *et al.*, eds., New York: Scientific American Books, pp. 567-581 (1992). An overview of suitable vectors or virions is provided in an article by Wilson, J.M. (*Clin. Exp. Immunol.* 107(Suppl. 1): 31-32, (1997)). Such vectors are derived from viruses that contain RNA (Vile, R.G., *et al.*, *Br. Med Bull.* 51(1): 12-30, (1995)) or DNA (Ali M., *et al.*, *Gene Ther.* 1(6): 367-384, (1994)). Example vector systems utilized in the art include the following: retroviruses (Vile, R.G., supra.), adenoviruses (Brody, S.L. *et al.*, *Ann. N.Y. Acad. Sci.* 716: 90-101, (1994)), adenoviral/retroviral chimeras (Bilbao, G., *et al.*, *FASEB J.* 11(8): 624-634, (1997)), adeno-associated viruses (Flotte, T.R. and Carter, B.J., *Gene Ther.* 2(6): 357-362, (1995)), herpes simplex virus (Latchman, D.S., *Mol. Biotechnol.* 2(2): 179-195, (1994)), Parvovirus (Shaughnessy, E., *et al.*, *Semin Oncol.* 23(1): 159-171, (1996)) and reticuloendotheliosis virus (Donburg, R., *Gene Therap.* 2(5): 301-310, (1995)). Also of interest in the art, the development of extrachromosomal replicating vectors for gene therapy (Calos, M.P., *Trends Genet.* 12(11): 463-466, (1996)).

Other, nonviral methods for gene transfer known in the art (Abdallah, B. *et al.*, *Biol. Cell* 85(1): 1-7, (1995)) might be utilized for the introduction of *de novo* DNA cytosine methyltransferase polynucleotides into target cells; for example, receptor-mediated DNA delivery (Philips, S.C., *Biologicals* 23(1): 13-16, (1995)) and lipidic vector systems (Lee, R.J. and Huang, L., *Crit. Rev. Ther. Drug Carrier Syst.* 14(2): 173-206, (1997)) are promising alternatives to viral-based delivery systems.

General methods for construction of gene therapy vectors and the introduction thereof into affected animals for therapeutic purposes may be obtained in the above-referenced publications, the disclosures of which are specifically incorporated herein by reference in their entirety. In one such general method, vectors comprising the isolated polynucleotides of the present invention are directly introduced into target cells or tissues of the affected animal, preferably by injection, inhalation, ingestion or introduction into a mucous membrane via solution; such an approach is generally referred to as "*in vivo*" gene therapy. Alternatively, cells, tissues or organs may be removed from the affected animal and placed into culture according to methods that are well-known to one of ordinary skill in the art; the vectors comprising the *de novo* DNA cytosine methyltransferase polynucleotides may then be introduced into these cells or tissues by any of the methods described generally above for introducing isolated polynucleotides into a cell or tissue, and, after a sufficient amount of time to allow incorporation of the *de novo* DNA cytosine methyltransferase polynucleotides, the cells or tissues may then be re-inserted into the affected animal. Since the introduction of a *de novo* DNA cytosine methyltransferase gene is performed outside of the body of the affected animal, this approach is generally referred to as "*ex vivo*" gene therapy.

For both *in vivo* and *ex vivo* gene therapy, the isolated *de novo* DNA cytosine methyltransferase polynucleotides of the invention may alternatively be operatively linked to a regulatory DNA sequence, which may be a *de novo* DNA cytosine methyltransferase promoter or an enhancer, or a heterologous regulatory

DNA sequence such as a promoter or enhancer derived from a different gene, cell or organism, to form a genetic construct as described above. This genetic construct may then be inserted into a vector, which is then used in a gene therapy protocol. The need for transcriptionally targeted and regulatable vectors providing cell-type specific and inducible promoters is well recognized in the art (Miller, N. and Whelan, J., *Hum. Gene Therap.* 8(7): 803-815, (1997); and Walther, W. and Stein, U., *Mol. Med. J.*, 74(7): 379-392, (1996)), and for the purposes of *de novo* DNA cytosine methyltransferase gene therapy, is incorporated herein by reference.

The construct/vector may be introduced into the animal by an *in vivo* gene therapy approach, *e.g.*, by direct injection into the target tissue, or into the cells or tissues of the affected animal in an *ex vivo* approach. In another preferred embodiment, the genetic construct of the invention may be introduced into the cells or tissues of the animal, either *in vivo* or *ex vivo*, in a molecular conjugate with a virus (*e.g.*, an adenovirus or an adeno-associated virus) or viral components (*e.g.*, viral capsid proteins; *see* WO 93/07283). Alternatively, transfected host cells, which may be homologous or heterologous, may be encapsulated within a semi-permeable barrier device and implanted into the affected animal, allowing passage of *de novo* DNA cytosine methyltransferase polypeptides into the tissues and circulation of the animal but preventing contact between the animal's immune system and the transfected cells (*see* WO 93/09222). These approaches result in increased production of *de novo* DNA cytosine methyltransferase by the treated animal via (a) random insertion of the *de novo* DNA cytosine methyltransferase gene into the host cell genome; or (b) incorporation of the *de novo* DNA cytosine methyltransferase gene into the nucleus of the cells where it may exist as an extrachromosomal genetic element. General descriptions of such methods and approaches to gene therapy may be found, for example, in U.S. Patent No. 5,578,461, WO 94/12650 and WO 93/09222.

Antisense oligonucleotides have been described as naturally occurring biological inhibitors of gene expression in both prokaryotes (Mizuno *et al.*, *Proc. Natl. Acad. Sci. USA 81*:1966-1970 (1984)) and eukaryotes (Heywood, *Nucleic Acids Res. 14*:6771-6772 (1986)), and these sequences presumably function by hybridizing to complementary mRNA sequences, resulting in hybridization arrest of translation (Paterson, *et al.*, *Proc. Natl. Acad. Sci. USA, 74*:4370-4374 (1987)).

Thus, another gene therapy approach utilizes antisense technology. Antisense oligonucleotides are short synthetic DNA or RNA nucleotide molecules formulated to be complementary to a specific gene or RNA message. Through the binding of these oligomers to a target DNA or mRNA sequence, transcription or translation of the gene can be selectively blocked and the disease process generated by that gene can be halted (*see*, for example, Jack Cohen, *Oligodeoxynucleotides, Antisense Inhibitors of Gene Expression*, CRC Press (1989)). The cytoplasmic location of mRNA provides a target considered to be readily accessible to antisense oligodeoxynucleotides entering the cell; hence much of the work in the field has focused on RNA as a target. Currently, the use of antisense oligodeoxynucleotides provides a useful tool for exploring regulation of gene expression *in vitro* and in tissue culture (Rothenberg, *et al.*, *J. Natl. Cancer Inst. 81*:1539-1544 (1989)).

Antisense therapy is the administration of exogenous oligonucleotides which bind to a target polynucleotide located within the cells. For example, antisense oligonucleotides may be administered systemically for anticancer therapy (Smith, International Application Publication No. WO 90/09180).

The antisense oligonucleotides of the present invention include derivatives such as S-oligonucleotides (phosphorothioate derivatives or S-oligos, *see*, Jack Cohen, *supra*). S-oligos (nucleoside phosphorothioates) are isoelectronic analogs of an oligonucleotide (O-oligo) in which a nonbridging oxygen atom of the phosphate group is replaced by a sulfur atom. The S-oligos of the present invention may be prepared by treatment of the corresponding O-oligos with 3*H*-1,2-benzodithiol-3-one-1,1-dioxide which is a sulfur transfer reagent. *See* Iyer *et*

al., J. Org. Chem. *55*:4693-4698 (1990); and Iyer *et al., J. Am. Chem. Soc.* *112*:1253-1254 (1990), the disclosures of which are fully incorporated by reference herein.

As described herein, sequence analysis of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3 or the SEQ ID NO:4 cDNA clone shows that sequence that is nonhomologous to known DNA methyltransferase sequences may be identified (*see* Figures 1 and 4). Thus, the antisense oligonucleotides of the present invention may be RNA or DNA that is complementary to and stably hybridize with such sequences that are specific for a *de novo* DNA cytosine methyltransferase gene of the invention. Use of an oligonucleotide complementary to such regions allows for selective hybridization to a *de novo* DNA cytosine methyltransferase mRNA and not to an mRNA encoding a maintenance methyltransferase protein.

Preferably, the antisense oligonucleotides of the present invention are a 15 to 30-mer fragment of the antisense DNA molecule coding for unique sequences of the *de novo* DNA cytosine methyltransferase cDNAs. Preferred antisense oligonucleotides bind to the 5′-end of the *de novo* DNA cytosine methyltransferase mRNAs. Such antisense oligonucleotides may be used to down regulate or inhibit expression of the gene.

Other criteria that are known in the art may be used to select the antisense oligonucleotides, varying the length or the annealing position in the targeted sequence.

Included as well in the present invention are pharmaceutical compositions comprising an effective amount of at least one of the antisense oligonucleotides of the invention in combination with a pharmaceutically acceptable carrier. In one embodiment, a single antisense oligonucleotide is utilized.

In another embodiment, two antisense oligonucleotides are utilized which are complementary to adjacent regions of the genome. Administration of two antisense oligonucleotides that are complementary to adjacent regions of the genome or corresponding mRNA may allow for more efficient inhibition of

genomic transcription or mRNA translation, resulting in more effective inhibition of protein or mRNA production.

Preferably, the antisense oligonucleotide is coadministered with an agent which enhances the uptake of the antisense molecule by the cells. For example, the antisense oligonucleotide may be combined with a lipophilic cationic compound which may be in the form of liposomes. The use of liposomes to introduce nucleotides into cells is taught, for example, in U.S. Patent Nos. 4,897,355 and 4,394,448, the disclosures of which are incorporated by reference in their entirety (*see also* U.S. Patent Nos. 4,235,871, 4,231,877, 4,224,179, 4,753,788, 4,673,567, 4,247,411, and 4,814,270 for general methods of preparing liposomes comprising biological materials).

Alternatively, the antisense oligonucleotide may be combined with a lipophilic carrier such as any one of a number of sterols including cholesterol, cholate and deoxycholic acid. A preferred sterol is cholesterol.

In addition, the antisense oligonucleotide may be conjugated to a peptide that is ingested by cells. Examples of useful peptides include peptide hormones, antigens or antibodies, and peptide toxins. By choosing a peptide that is selectively taken up by the targeted tissue or cells, specific delivery of the antisense agent may be effected. The antisense oligonucleotide may be covalently bound via the 5'OH group by formation of an activated aminoalkyl derivative. The peptide of choice may then be covalently attached to the activated antisense oligonucleotide via an amino and sulfhydryl reactive hetero bifunctional reagent. The latter is bound to a cysteine residue present in the peptide. Upon exposure of cells to the antisense oligonucleotide bound to the peptide, the peptidyl antisense agent is endocytosed and the antisense oligonucleotide binds to the target mRNA to inhibit translation (Haralambid *et al.*, WO 8903849 and Lebleu *et al.*, EP 0263740).

The antisense oligonucleotides and the pharmaceutical compositions of the present invention may be administered by any means that achieve their intended purpose. For example, administration may be by parenteral,

subcutaneous, intravenous, intramuscular, intraperitoneal, or transdermal routes. The dosage administered will be dependent upon the age, health, and weight of the recipient, kind of concurrent treatment, if any, frequency of treatment, and the nature of the effect desired.

Compositions within the scope of this invention include all compositions wherein the antisense oligonucleotide is contained in an amount effective to achieve the desired effect, for example, inhibition of proliferation and/or stimulation of differentiation of the subject cancer cells. While individual needs vary, determination of optimal ranges of effective amounts of each component is with the skill of the art.

Alternatively, antisense oligonucleotides can be prepared which are designed to interfere with transcription of the gene by binding transcribed regions of duplex DNA (including introns, exons, or both) and forming triple helices (e.g., *see* Froehler *et al.*, WO 91/06626 or Toole, WO 92/10590). Preferred oligonucleotides for triple helix formation are oligonucleotides which have inverted polarities for at least two regions of the oligonucleotide (*Id.*). Such oligonucleotides comprise tandem sequences of opposite polarity such as 3′---5′-L-5′---3′, or 5′---3′-L-3′---5′, wherein L represents a 0-10 base oligonucleotide linkage between oligonucleotides. The inverted polarity form stabilizes single-stranded oligonucleotides to exonuclease degradation (Froehler *et al., supra*). The criteria for selecting such inverted polarity oligonucleotides is known in the art, and such preferred triple helix-forming oligonucleotides of the invention are based upon SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3 or SEQ ID NO:4.

In therapeutic application, the triple helix-forming oligonucleotides can be formulated in pharmaceutical preparations for a variety of modes of administration, including systemic or localized administration, as described above.

The antisense oligonucleotides of the present invention may be prepared according to any of the methods that are well known to those of ordinary skill in the art, as described above.

Ribozymes provide an alternative method to inhibit mRNA function. Ribozymes may be RNA enzymes, self-splicing RNAs, and self-cleaving RNAs (Cech *et al.*, *Journal of Biological Chemistry 267*:17479-17482 (1992)). It is possible to construct *de novo* ribozymes which have an endonuclease activity directed in *trans* to a certain target sequence. Since these ribozymes can act on various sequences, ribozymes can be designed for virtually any RNA substrate. Thus, ribozymes are very flexible tools for inhibiting the expression of specific genes and provide an alternative to antisense constructs.

A ribozyme against chloramphenicol acetyltransferase mRNA has been successfully constructed (Haseloff *et al.*, *Nature 334*:585-591 (1988); Uhlenbeck *et al.*, *Nature 328*:596-600 (1987)). The ribozyme contains three structural domains: 1) a highly conserved region of nucleotides which flank the cleavage site in the 5′ direction; 2) the highly conserved sequences contained in naturally occurring cleavage domains of ribozymes, forming a base-paired stem; and 3) the regions which flank the cleavage site on both sides and ensure the exact arrangement of the ribozyme in relation to the cleavage site and the cohesion of the substrate and enzyme. RNA enzymes constructed according to this model have already proved suitable *in vitro* for the specific cleaving of RNA sequences (Haseloff *et al.*, *supra*).

Alternatively, hairpin ribozymes may be used in which the active site is derived from the minus strand of the satellite RNA of tobacco ring spot virus (Hampel *et al.*, *Biochemistry 28*:4929-4933 (1989)). Recently, a hairpin ribozyme was designed which cleaves human immunodeficiency virus type 1 RNA (Ojwang *et al.*, *Proc. Natl. Acad. Sci. USA 89*:10802-10806 (1992)). Other self-cleaving RNA activities are associated with hepatitis delta virus (Kuo *et al.*, *J. Virol. 62*:4429-4444 (1988)).

As discussed above, preferred targets for ribozymes are the *de novo* DNA cytosine methyltransferase nucleotide sequences that are not homologous with maintenance methyltransferase sequences such as Dnmt 1 or Dnmt 2. Preferably, the ribozyme molecule of the present invention is designed based upon the

chloramphenicol acetyltransferase ribozyme or hairpin ribozymes, described above. Alternatively, ribozyme molecules are designed as described by Eckstein *et al.* (International Publication No. WO 92/07065) who disclose catalytically active ribozyme constructions which have increased stability against chemical and enzymatic degradation, and thus are useful as therapeutic agents.

In an alternative approach, an external guide sequence (EGS) can be constructed for directing the endogenous ribozyme, RNase P, to intracellular mRNA, which is subsequently cleaved by the cellular ribozyme (Altman *et al.*, U.S. Patent No. 5,168,053). Preferably, the EGS comprises a ten to fifteen nucleotide sequence complementary to an mRNA and a 3'-NCCA nucleotide sequence, wherein N is preferably a purine (*Id.*). After EGS molecules are delivered to cells, as described below, the molecules bind to the targeted mRNA species by forming base pairs between the mRNA and the complementary EGS sequences, thus promoting cleavage of mRNA by RNase P at the nucleotide at the 5'side of the base-paired region (*Id.*).

Included as well in the present invention are pharmaceutical compositions comprising an effective amount of at least one ribozyme or EGS of the invention in combination with a pharmaceutically acceptable carrier. Preferably, the ribozyme or EGS is coadministered with an agent which enhances the uptake of the ribozyme or EGS molecule by the cells. For example, the ribozyme or EGS may be combined with a lipophilic cationic compound which may be in the form of liposomes, as described above. Alternatively, the ribozyme or EGS may be combined with a lipophilic carrier such as any one of a number of sterols including cholesterol, cholate and deoxycholic acid. A preferred sterol is cholesterol.

The ribozyme or EGS, and the pharmaceutical compositions of the present invention may be administered by any means that achieve their intended purpose. For example, administration may be by parenteral, subcutaneous, intravenous, intramuscular, intra-peritoneal, or transdermal routes. The dosage administered will be dependent upon the age, health, and weight of the recipient,

kind of concurrent treatment, if any, frequency of treatment, and the nature of the effect desired. For example, as much as 700 milligrams of antisense oligodeoxynucleotide has been administered intravenously to a patient over a course of 10 days (i.e., 0.05 mg/kg/hour) without signs of toxicity (Sterling, "Systemic Antisense Treatment Reported," *Genetic Engineering News 12*(12):1, 28 (1992)).

Compositions within the scope of this invention include all compositions wherein the ribozyme or EGS is contained in an amount which is effective to achieve inhibition of proliferation and/or stimulate differentiation of the subject cancer cells, or alleviate AD. While individual needs vary, determination of optimal ranges of effective amounts of each component is with the skill of the art.

In addition to administering the antisense oligonucleotides, ribozymes, or EGS as a raw chemical in solution, the therapeutic molecules may be administered as part of a pharmaceutical preparation containing suitable pharmaceutically acceptable carriers comprising excipients and auxiliaries which facilitate processing of the antisense oligonucleotide, ribozyme, or EGS into preparations which can be used pharmaceutically.

Suitable formulations for parenteral administration include aqueous solutions of the antisense oligonucleotides, ribozymes, EGS in water-soluble form, for example, water-soluble salts. In addition, suspensions of the active compounds as appropriate oily injection suspensions may be administered. Suitable lipophilic solvents or vehicles include fatty oils, for example, sesame oil, or synthetic fatty acid esters, for example, ethyl oleate or triglycerides. Aqueous injection suspensions may contain substances which increase the viscosity of the suspension include, for example, sodium carboxymethyl cellulose, sorbitol, and/or dextran. Optionally, the suspension may also contain stabilizers.

Alternatively, antisense RNA molecules, ribozymes, and EGS can be coded by DNA constructs which are administered in the form of virions, which are preferably incapable of replicating *in vivo* (*see*, for example, Taylor, WO 92/06693). For example, such DNA constructs may be administered using

herpes-based viruses (Gage *et al.*, U.S. Patent No. 5,082,670). Alternatively, antisense RNA sequences, ribozymes, and EGS can be coded by RNA constructs which are administered in the form of virions, such as retroviruses. The preparation of retroviral vectors is well known in the art (*see*, for example, Brown *et al.*, "Retroviral Vectors," in *DNA Cloning: A Practical Approach*, Volume 3, IRL Press, Washington, D.C. (1987)).

Specificity for gene expression may be conferred by using appropriate cell-specific regulatory sequences, such as cell-specific enhancers and promoters. Such regulatory elements are known in the art, and their use enables therapies designed to target specific tissues, such as liver, lung, prostate, kidney, pancreas, etc., or cell populations, such as lymphocytes, neurons, mesenchymal, epithelial, muscle, etc.

In addition to the above noted methods for inhibiting the expression of the de novo methyltransferase genes of the invention, gene therapeutic applications may be employed to provide expression of the polypeptides of the invention.

Examples

Example 1: Cloning and Sequence Analysis of the Mouse Dnmt3a and Dnmt3b and the Human DNMT3A and DNMT3B Genes and Polypeptides

In search of a mammalian *de novo* DNA methyltransferase, two independent approaches were undertaken, based on the assumption that an unknown mammalian DNA methyltransferase must contain the highly conserved cytosine methyltransferase motifs in the catalytic domain of known methyltransferases (Lauster, R. *et al.*, *J. Mol. Biol. 206*:305-312 (1989) and Kumar, S. *et al.*, *Nucl. Acids Res. 22*:1-10 (1994)). Our first approach, an RT/PCR-based screening using oligonucleotide primers corresponding to the conserved motifs of the known cytosine DNA methyltransferases, failed to detect any novel methyltransferase gene from *Dnmt1* null ES cells (data not shown).

The second approach was a tblastn search of the dbEST database using full length bacterial cytosine methyltransferase sequences as queries.

A search of the dbEST database was performed with the tblastn program (Altschul, S. F. *et al., J. Mol. Biol. 215*:403-410 (1990)) using bacterial cytosine methyltransferases as queries. Candidate EST sequences were used one by one as queries to search the non-redundant protein sequence database in GenBank with the blastx program. This process would eliminate EST clones corresponding to known genes (including known DNA methyltransferases) and those which show a higher similarity to other sequences than to DNA methyltransferases. Two EST clones (GenBank numbers W76111 and N88352) were found after the initial search. Two more EST sequences (fl2227 and T66356) were later found after a blastn search of dbEST with the EST sequence of W76111 as a query. Two of the EST clones (W76111 and T66356) were deposited by the I.M.A.G.E. Consortium (Lawrence Livermore National Laboratory, Livermore, CA) and obtained from American Type Culture Collection (Manassas, VA). Sequencing of these two cDNA clones revealed that they were partial cDNA clones with large open reading frames corresponding to two related genes. The translated amino acid sequences revealed the presence of the highly conserved motifs characteristic of DNA cytosine methyltransferases. The EST sequences were then used as probes for screening mouse E7.5 embryo and ES cell cDNA libraries and a human heart cDNA library (Clontech, CA).

In a screening of the dbEST database using 35 bacterial cytosine-5 DNA methyltransferase sequences as queries, eight EST clones were found to have the highest similarity but not to be identical to the known cytosine-5-DNA methyltransferase genes. Six of the eight EST sequences were deposited by the I.M.A.G.E. Consortium (Lawrence Livermore National Laboratory, Livermore, CA) and obtained from TIGR/ATCC (American Type Culture Collection, Manassas, VA). Sequencing of these 6 cDNA clones revealed that they were partial cDNA clones with large open reading frames corresponding to three novel genes. The translated amino acid sequences revealed the presence of the highly

conserved motifs characteristic of DNA cytosine methyltransferases. The EST sequences were then used as probes for screening a mouse ES cell cDNA library, a mouse E11.5 embryonic cDNA library (Clontech, CA) and human heart cDNA library.

Human and mouse cDNA libraries were screened using EST sequences as probes. Sequencing analysis of several independent cDNA clones revealed that two homologous genes were present in both human and mouse. This was further confirmed by Southern analysis of genomic DNA, intron/exon mapping and sequencing of genomic DNA (data not shown). The full length mouse cDNAs for each gene were assembled and complete sequencing revealed that both genes contained the highly conserved cytosine methyltransferase motifs and shared overall 51% of amino acid identity (76% identity in the catalytic domain) (Fig. 3). Since these two genes showed little sequence similarities to Dnmt1(Bestor, T. H. *et al., J. Mol. Biol. 203*:971-983 (1988) and Yen, R-W. C. *et al., Nucleic Acids Res. 20*:2287-2291 (1992)) and a recently cloned putative DNA methyltransferase gene, Dnmt2 (*see* Yoder, J. A. and Bestor, T. H. *Hum. Mol. Genet.* 7:279-284 (1998)) and Okano, M., Xie, S. and Li, E., (submitted)), beyond the conserved methyltransferase motifs in the catalytic domain, they were named Dnmt3a and Dnmt3b.

The full length Dnmt3a and Dnmt3b genes encode 908 and 859 amino acid polypeptides, termed Dnmt3a and Dnmt3b1, respectively. Nucleotide and amino acid sequences of each are presented in Figures 1A, 1B, 2A, and 2B. The Dnmt3b gene also produces through alternative splicing at least two shorter isoforms of 840 and 777 amino acid residues, termed Dnmt3b2 and Dnmt3b3, respectively, (Fig. 4).

To obtain full length human cDNA, fetal heart and fetal testis cDNA libraries were screened using EST clones as probes. Sequencing analysis of several overlapping DNMT3A cDNA clones indicates that the DNMT3A gene encodes a polypeptide of 912 amino acid residues. DNMT3B cDNA clones were not detected in the fetal heart library, but several DNMT3B cDNA clones were

obtained after screening the fetal testis library. PCR screening of large cDNA clones from 24 human tissues was also performed using the Human Rapid-Screen™ cDNA Library Panels (OriGene Technologies, MD). The largest cDNA clone contained a 4.2 kb insert from a small intestine cDNA library. Sequencing analysis of overlapping cDNA clones indicated that the deduced full length DMNT3B consists of 853 amino acid residues. Since in-frame stop codons are found upstream of the ATG of both DNMT3A and DNMT3B, it is concluded that these cDNA clones encode full-length DNMT3A and DNMT3B proteins.

The full length human DNMT3A and DNMT3B cDNAs encode 912 and 853 amino acid polypeptides, termed DNMT3A and DNMT3B1, respectively. Nucleotide and polypeptide sequences are presented in Figures 1C, 1D, 2C and 2D, respectively. The DNMT3B gene also produces through alternative splicing at least two shorter isoforms, termed DNMT3B2 and DNMT3B3, respectively. DNMT3B2 comprises amino acid residues 1 to 355 and 376 to 853 of SEQ ID NO:4; and DNMT3B3 comprises amino acid residues 1 to 355 and 376 to 743 and 807 to 853 of SEQ ID NO:4.

Also identified through screening was a related zebrafish gene, termed Zmt-3, which from the EST database (GenBank number AF135438).

The GenBank STS database was used to map chromosome localization by using DNMT3A and DNMT3B sequences as queries. The results identified markers WI-6283 (GenBank Accession number G06200) and SHGC-15969 (GenBank Accession number G15302), which matched the cDNA sequence of DNMT3A and DNMT3B, respectively. WI-6283 has been mapped to 2p23 between D2S171 and D2S174 (48-50 cM) on the radiation hybrid map by Whitehead Institute/MIT Center for Genome Research. The corresponding mouse chromosome location is at 4.0 cM on chromosome 12. SHGC-15969 has been mapped to 20pl 1.2 between D20S184 and D20S106 (48-50 cM) by Stanford Human Genome Center. The corresponding mouse chromosome locus is at 84.0 cM on chromosome 2.

Taking the advantage of the newly identified DNMT3A and DNMT3B cDNA sequences, the human genomic sequence database was searched by BLAST. While human DNMT3A cDNA did not match any related genornic sequences in the database, a DNMT3B genomic YAC clone from GenBank (AL035071) was identified when DNMT3B cDNA sequences were used as queries.

The DNMT3B cDNA and the genomic DNA GenBank (AL035071) clone were used to map all exons using BESTFIT of the GCG program. As shown in Figure 4C, there are total 23 exons, spanning some 48 kb genomic DNA. The putative first exon is located within a CpG island where the promoter is probably located as predicted by the GENSCAN program (Whitehead/MIT Center for Genome Research).

Sequencing of various cDNA clones indicates that the human DNMT3B gene contains three alternatively spliced exons, exons 10, 21 and 22. Similar to the mouse gene, DNMT3B1 contains all 23 exons, whereas DNMT3B2 lacks exon 10 and DNMT3B3 lacks exons 10, 21 and 22. The nucleotide sequences at the exon/intron boundaries are shown in Figure 4D. The elucidation of human *DNMT3B* gene structure may facilitate analysis of DNMT3B mutations in certain cancers with characteristic hypomethylation of genomic: DNA (Narayan, A., *et al.*, *Int. J. Cancer 77*:833-838 (1998); Qu, G., *et al.*, *Mutan. Res. 423*:91-101 (1999)).

Figure 3A presents an alignment of mouse Dnmt3a and Dnmt3b polypeptide sequences that was accomplished using the GCG program. The vertical lines indicate amino acid identity, while the dots and the colons indicate similarities. Dots in amino acid sequences indicate gaps introduced to maximize alignment. The conserved Cys-rich region is shaded. The full length mouse Dnmt3a and Dnmt3b genes encode 908 and 859 amino acid polypeptides. Furthermore, the analysis reveals that both genes contained the highly conserved cytosine methyltransferase motifs and share overall 51% of amino acid identity (76% identity in the catalytic domain). The Dnmt3b gene also produces at least two shorter isoforms of 840 and 777 amino acid residues, termed Dnmt3b2 and Dnmt3b3, respectively, through alternative splicing (Fig. 4).

Figure 3B presents a GCG program alignment using the of the protein sequences of human DNMT3A and DNMT3B1. Vertical lines represent identical amino acid residues, whereas dots represent conserved changes. Dots in amino acid sequences indicate gaps introduced to maximize alignment.

In Figure 4A, presents a schematic diagram of the overall protein structures for mouse *Dnmt*1, mouse *Dnmt*2, a putative methyltransferase, and the family of Dnmt3a and Dnmt3b(1-3) methyltransferases. *Dnmt*1, Dnmt3a and Dnmt3bs all have a putative N-terminal regulatory domain. The filled bars represent the five conserved methyltransferase motifs (I, IV, VI, IX, and X). The shaded boxes in Dnmt3a and Dnmt3bs represent the Cys-rich region that shows no sequence homology to the Cys-rich, Zn^{2+}-binding region of *Dnmt*1 polypeptide. Sites of alternative splicing at amino acid residues 362-383 and 749-813 in Dnmt3bs are indicated.

An analysis of the human DNMT3 proteins provides similar results as with the mouse Dnmt proteins. Figure 4B presents a similar schematic of the human DNMT3 proteins and zebrafish Znmt3 protein. The homology between differences between these DNMT3 proteins is indicated by the percentage of sequence identity when compared to DNMT3A.

In addition, the genomic organization of the human DNMT3B1 locus is presented in Figure 4C as possessing 23 exons (filled rectangles), a CpG island (dotted rectangle),a translation initiation codon (ATG) and a stop codon (TAG) in exons 2 and 23, respectively. Figure 4D presents the size of the exons and introns as well as sequences (uppercase for exons and lowercase for introns) at exon/intron boundaries.

In Figure 5, sequence analysis of the catalytic domain indicates that this new family of DNA methyltransferases contains conserved amino acid residues in each of the five highly conserved motifs, but significant differences are discernible when compared to the known consensus sequences.

Figure 5A presents an alignment by ClustalW 1.7 of the amino acid sequences of the five highly conserved motifs in eukaryotic methyltransferase

genes. Amino acid residues which are conserved in five or more genes are highlighted. The Dnmt3 family methyltransferases are most closely related to a bacterial DNA methyltransferase (*M. Spr.*). Sequence comparison of the catalytic domain of all known eukaryotic DNA methyltransferases and most of the bacterial cytosine methyltransferases used in the tblastn search indicates that this family of methyltransferases are distantly related to all the known eukaryotic DNA methyltransferases, including the *Dnmt* 1 polypeptide from vertebrate and plant (Bestor, T. H. *et al., J. Mol. Biol. 203*:971-983 (1988), Yen, R-W. C. *et al., Nucleic Acids Res. 20*:2287-2291 (1992) and Finnegan, E. J. and Dennis, E. S. *Nucleic Acids Res. 21*:2383-2388 (1993)); the human and mouse *Dnmt* 2 polypeptides (Yoder, J. A. and Bestor, T. H. *Hum. Mol. Genet. 7*:279-284 (1998), Okano, M., Xie, S. & Li, E., (submitted)); and masc1 from Ascobolus (Malagnac, F. *et al., Cell 91*:281-290 (1997)), indicating that the Dnmt3 gene family originated from a unique prokaryotic prototype DNA methyltransferase during evolution.

The cysteine-rich region located upstream of the catalytic domain was found to be conserved among all of the DNMT3 proteins (Fig. 5B). This Cysteine-rich region, however, is unrelated to the Cysteine-rich (or Zn^{2+} -binding) region of DNMT1 (Bestor, T.H., *et al., J. Mo. Biol. 203*:971-983 (1998); Bestor, T.H., *EMBO J. 11*:2611-2617 (1992)). Interestingly, the Cysteine-rich domain of DNMT3 proteins shares homology with a similar domain found in the X-linked ATRX gene of the SNF2/SWI family (Picketts, D.J., *et al., Hum. Mol. Genet. 5*:1899-1907 (1996)), raising the interesting possibility that this domain may mediate protein-protein or protein-DNA interactions.

The evolutionary relatedness of cytosine-5 methyltransferases as shown by a non-rooted phylogenic tree is presented in Figure 5C. Amino acid sequences from motif I to motif VI of bacterial and eukaryotic cytosine-5 methyltransferases were used for sequence alignment, and the alignment data was analyzed by ClustalW 1.7 under conditions excluding positions with gaps. Results were visualized utilizing Phlip version 3.3. Amino acid sequences from motif IX to

motif X were also analyzed and provided similar results (data not shown). (Abbreviation Ath; *Arabidopsis thaliana*, Urc; sea urchin, Xen; *Xenopus laevis*).

Example 2: Baculovirus-mediated Expression of Dnmt3a and Dnmt3b

To test whether the newly cloned Dnmt3 genes encode active DNA methyltransferases, the cDNAs of Dnmt3a, Dnmt3b1, Dnmt3b2, and Dnmt1 were overexpressed in insect cells using the baculovirus-mediated expression system (Clontech, CA).

To construct the Dnmt3a expression vector, pSX134, the *Xma I/Eco RI* fragment of Dnmt3a cDNA was first cloned into the *Nco I/Eco RI* sites of pET2 ld with the addition of an *Xma I/Nco I* adapter (SX165: 5'-CATGGGCAGCAGCCATCATCATCATCATCATGGGAATTCCATGCCC TCCAGCGGCC and SX166: 5'-CCGGGGCCGCTGGAGGGCATGGA ATTCCCATGATGATGATGATGATGGCTGCTGCC) that produced pSX132His. pSX134 was obtained by cloning the *EcoR I/Xba I* fragment of pSX 132His into the *EcoR I/Xba I* sites of pBacPAK9. The Dnmt3b1 and Dnmt3b2 expression vectors, pSX153 and pSX154, were constructed by cloning *Eco RI* fragments of Dnmt3b1 and Dnmt3b2 cDNA into the *Eco RI* site of pBacPAK9, respectively. The Dnmt1 expression vector pSX148 was constructed by cloning the *Bgl I/Sac I* fragment of Dnmt1 cDNA into the *Bgl II/Sac I* sites of pBacPAK-His2 with the addition of a *Bgl I/Bgl II* adapter (SX180: 5'-GATCTATGCCAGCGCGAACAGCTCCAGCCCGAGTGCCTGCGCTTGC CTCCC and SX181: 5'- AGGCAAGCGCAGGCACTCGGGCTGGAGCTGTT CGCGCTGGCATA).

pSX134 (Dnmt3a), pSX153 (Dnmt3b1), pSX153 (Dnmt3b2) and pSX148 (*Dnmt1*) were used to make the recombinant baculoviruses according to the procedures recommended by the manufacturer. T175 flasks were used for cell culture and virus infection. Sf21 host cells were grown in the SF-900 II SFM medium with 10% of the certified FBS (both from GIBCO, MD) and infected

with the recombinant viruses 12-24 hours after the cells were split when they reached 90-95% affluence. After 3 days, the infected insect cells were harvested and frozen in the liquid nitrogen for future use.

Example 3: RNA Expression Analysis

ES cells were routinely cultured on a feeder layer of mouse embryonic fibroblasts in DMEM medium containing LIF (500 units/ml) and were differentiated as embryoid bodies in suspension culture as described (Lei, H., *et al., Development 122*:3195-3205 (1996)). Ten days after seeding, embryoid bodies were harvested for RNA preparation.

Total RNA was prepared from ES cells, ovary and testis tissue using the GTC-CsCl centrifugation method, fractionated on a formaldehyde denaturing 1% agarose gel by electrophoresis and transferred to a nylon membrane. PolyA+ RNA blots (2μg per lane) of mouse and human tissues were obtained from Clontech, CA. All blots were hybridized to random-primed cDNA probes in hybridization solution containing 50% formamide at 42°C and washed with 0.2 X SSC, 0.1% SDS at 65°C and exposed to X-ray film (Kodak).

Fig. 6A presents mouse polyA+ RNA blots of adult tissues (left) and embryos (right) probed with full length Dnmt3a, Dnmt3b and a control β-actin cDNA probe. Each lane contains 2 μg of polyA+ RNA. (Ht, Heart; Br, Brain; Sp, Spleen; Lu, Lung; Li, Liver; Mu, Skeletal Muscle; Ki, Kidney; Te, Testis; and embryos at gestation days 7 (E7), 11 (E11), 15 (E15), and 17 (E17). Fig. 6B is a mouse total RNA blot (10 μg per lane) of ES cell and adult organ RNA samples and Fig. 6C shows a mouse total RNA blot (20 μg per lane) of undifferentiated (Undiff.) and differentiated (Diff.) ES cells RNA hybridized to Dnmt3a, Dnmt3b or β-actin probes.

It has been shown that the maintenance methylation activity is constitutively present in proliferating cells, whereas the *de novo* methylation activity is highly regulated. Active *de novo* methylation has been shown to occur

primarily in ES cells (or embryonic carcinoma cells), early postimplantation embryos and primordial germ cells (Jähaner, D. and Jaenish, R., "DNA Methylation in Early Mammalian Development," In *DNA Methylation: Biochemistry and Biological Significance*, Razin, A. *et al., eds.*, Springer-Verlag (1984) pp. 189-219; Razin, A., and Cedar, H., "DNA Methylation and Embryogenesis," in *DNA Methylation: Molecular Biology and Biological Significance*, Jost., J. P. *et al., eds.*, Birkhäuser Verlag, Basel, Switzerland (1993) pp. 343-357; Chaillet, J. R. *et al., Cell 66*:77-83 (1991); and Li, E. "Role of DNA Methylation in Development," in *Genomic Imprinting: Frontiers in Molecular Biology*, Reik, W. and Sorani, A. *eds.*, IRL Press, Oxford (1997) pp. 1-20). The expression of both Dnmt3a and Dnmt3b in mouse embryos, adult tissues and ES cells was examined. The results indicate that two Dnmt3a transcripts, 9.5 kb and 4.2kb, are present in embryonic and adult tissue RNA. The 4.2 kb transcript, corresponding to the size of the full length cDNA, was expressed at very low levels in most tissues, except for the E11.5 embryo sample (Fig. 6A). A single 4.4 kb Dnmt3b transcript is detected in embryo and adult organ RNAs, with relatively high levels in testes and E11.5 embryo samples (Fig. 6A). Interestingly, both genes are expressed at much higher levels in ES cells than in adult tissues (Fig. 6B), and their expression decreased dramatically upon differentiation of ES cells in culture (Fig. 6C). In addition, Dnmt3a and Dnmt3b expression levels are unaltered in Dnmt1-deficient ES cells (Fig. 6C), suggesting that regulation of Dnmt3a and Dnmt3b expression is independent of Dnmt1.

These results suggest that both Dnmt3a and Dnmt3b are expressed specifically in ES cells and E11.5 embryo and/or testes. The expression in the E11.5 embryo and testes may correlate with the presence of developing or mature germ cells in these tissues. Therefore, the expression pattern of Dnmt3a and Dnmt3b appears to correlate well with *de novo* methylation activities in development.

For the RNA expression analysis of human DNMT3 genes, polyA+ RNA blots were hybridized using DNMT3A and DNMT3B cDNA fragments as probes.

Results indicate that DNMT3A RNA was expressed ubiquitously and was readily detected in most tissues examined at levels slightly lower than DNMT1 RNA (Fig.**X**). Three major DNMT3A transcripts, approximately 4.0, 4.4, and 9.5 kb, were detected. The relative expression level of the transcripts appeared to vary from tissue to tissue. Transcripts of similar sizes were also detected in mouse tissues. Results utilizing DNMT3B cDNA probes indicate that transcripts of about 4.2 kb were expressed at much lower levels in most tissues, but could be readily detected in the testis, thyroid and bone marrow (Fig. 9). Sequence analyses of different cDNA clones indicate the presence of alternatively spliced transcripts, although the size differences between these transcripts are too small to be detected by Northern analysis.

Hypermethylation of tumor suppressor genes is a common epigenetic lesion found in tumor cells (Laird, P.W. & Jaenisch, R., Ann. Rev. Genet. 30:441-464 (1996); Baylin, S.B., Adv. Cancer Res. 72:141-196 (1998)). To investigate whether DNMT3A and DNMT38 am abnormally activated in tumor cells, DNMT3 RNA expression was analyzed in several tumor cell lines by Northern blot hybridization. Results demonstrated that DNMT3A was expressed at higher levels in most tumor cell lines examined. (Figure 10). As in the normal tissues, three different size transcripts were also detected in tumor cells. The ratio of these transcripts appeared to be variable in different tumor cell lines. DNMT3B expression was dramatically elevated in most tumor cell lines examined though it was expressed at very low levels in normal adult tissues (Figure 10). The expression levels of both DNMT3A and DNMT3B appear to be comparable and proportional to that of DNMT1.

The murine Dnmt3a and Dnmt3b genes are highly expressed in undifferentiated ES cells, consistent with their potential role in *de novo* methylation during early embryonic development. Additionally, both genes are highly expressed in early embryos. Differences in their expression patterns in adult tissues in both human and mice suggest that each gene may have a distinct function in somatic tissues and may methylate different genes or genomic

sequences. The elevated expression of DNMT3 genes in human tumor cell lines suggests that the DNMT3 enzyme may be responsible for *de novo* methylation of CpG islands in tumor suppressor genes during tumor formation.

Example 4: Methyltransferase Activity Assay

In order to demonstrate DNA cytosine methyltransferase activity, the polypeptides of the invention were expressed and purified from recombinant host cells for use in *in vitro* assays.

Infected insect Sf21 cells and NIH3T3 cells were homogenized by ultrasonication in lysis solution (20 mM Tris-HCl, pH7.4, 10 mM EDTA, 500 mM NaCl, 10% glycerol, lmM DTT, lmM PMSF, 1 ug/ml leupeptin, 10 ug/ml TPCK, 10 ug/ml TLCK) and cleared by centrifugation at 100,000 g for 20 min.

The methyltransferase enzyme assay was carried out as described previously (Lei, H. *et al., Development 122*:3195-3205 (1996)). DNA substrates used in the assays include: poly (dI-dC), poly (dG-dC) (Pharmacia Biotech), lambda phage DNA (Sigma), pBluescriptIISK (Stratagene, CA), pMu3 plasmid, which contains tandem repeats of 535bp RsaI-RsaI fragment of MMLV LTR region in pUC9, and oligonucleotides. The oligonucleotide sequences utilized include:
#1, 5'-AGACMGGTGCCAGMGCAGCTGAGCMGGATC-3',
#2, 5'-GATCMGGCTCAGCTGMGCTGGCACMGGTCT-3',
#3, 5'-AGACCGGTGCCAGCGCAGCTGAGCCGGATC-3', and
#4, 5'-GATCCGGCTCAGCTGCGCTGGCACCGGTCT-3' (M represents 5-methylcytosine).

These sequences are the same as described in a previous study (Pradhan, S. *et al., Nucleic Acids Res. 25*:4666-4673 (1997)). Oligonucleotides were synthesized and purified by polyacrylamide gel electrophoresis (PAGE). To make double strand oligonucleotides, equimolar amounts of the two complimentary oligonucleotides were heated at 94°C for 10 min., mixed,

incubated at 78°C for 1 hr and cooled down slowly at room temperature. The annealing products were quantified for the yield of double-stranded oligonucleotides (dsDNA) by PAGE and methylene blue staining. In all cases, the yield of dsDNA was higher than 95%. The dsDNA of #1 and #2 were used as 'fully' methylated substrates, dsDNA of #1 and #4 as the hemi-methylated substrates, and dsDNA of #3 and #4 as unmethylated substrates.

For Southern analysis of the methylation of retrovirus DNA, 2 ug of pMMLV8.3, an 8.3kb *Hind* III fragment of Moloney murine leukemia virus cDNA in pBluescriptIISK, was methylated *in vitro* for 15 hrs under the same reaction conditions described above except that 160 uM of cold SAM was used instead of ^{3}H-methyl SAM. Then, an equal volume of the solution containing 1% SDS, 400 mM NaCl, and 0.2 mg/ml Proteinase K was added, and the sample was incubated at 37°C for 1 hr. After phenol/chloroform extraction, DNA was precipitated with ethanol, dried and dissolved in TE buffer. This procedure was repeated 5 times. An aliquot of DNA was purified after the first, third and fifth reaction, digested with *Hpa* II or *Msp* I in combination with *Kpn* I for 16 hrs, separated on 1% agarose gels, blotted and hybridized to the pMu3 probe.

In a standard methyltransferase assay, enzyme activity was detected with protein extracts from Sf21 cells overexpressing Dnmt3a and Dnmt3b polypeptides. Similar to the results obtained with the Dnmt1 polypeptide, the overexpressed Dnmt3 proteins were able to methylate various native and synthetic DNA substrates, among which poly(dI-dC) consistently gave rise to the highest initial velocity (Fig.7a). An analysis of the methylation of *Hpa* II sites in retroviral DNA by these enzymes was also performed. An MMLV full length cDNA was methylated for 1-5 times by incubation with protein extract from control Sf21 cells or Sf21 cells infected with baculoviruses expressing *Dnmt*1, Dnmt3a or Dnmt3b polypeptides. The *Hpa* II/*Msp* I target sequence, CCGG, is resistant to the *Hpa* II restriction enzyme, but sensitive to *Msp* I digestion when the internal C is methylated, and the restriction site becomes resistant to *Msp* I digestion when the external C is methylated (Jentsch, S. *et al.*, *Nucleic Acids Res.*

9:2753-2759 (1981)). Both Dnmt3a and Dnmt3b polypeptides could methylate multiple *Hpa* II sites in the 3' LTR regions of the MMLV DNA, as indicated by the presence of *Hpa* II-resistant fragments, though less efficiently than *Dnmt*1 polypeptide (Fig. 7b). Significantly, even after five consecutive rounds of *in vitro* methylation, the viral DNA was completely digested by *Msp* I. This result indicates that both Dnmt3a and Dnmt3b polypeptides methylate predominantly the internal cytosine residues, therefore, CpGs. Previously it was shown that the same region of the proviral DNA was efficiently methylated in *Dnmt1* null ES cells infected by the MMLV virus (Lei, H. *et al., Development 122*:3195-3205 (1996)).

Fig. 7A shows ^{3}H-methyl incorporation into different DNA substrates (poly (dI-dC), poly (dG-dC) (squares), lambda phage DNA (circles), pBluescriptIISK (triangles), and pMu3 (diamonds)) when incubated with protein extracts of Sf21 cells expressing Dnmt1, Dnmt3a, or Dnmt3b1. Fig. 7B shows Southern blot analysis of the *in vitro* methylation of untreated pMMLV DNA (lanes 1-3) and pMMLV DNA incubated with MT1 (lane 4-10), MT3α (lanes 11-15), MT3β (lanes 16-20) or control Sf21 (lanes 21-25) extracts that were digested with *Kpn* I(K), *Kpn* I and *Msp* I (K/M) or *Kpn* I and *Hpa* II (K/H). Restriction enzyme digested samples were then subjected to Southern blot analysis using the pMu3 probe.

*Dnmt*1 protein appears to function primarily as a maintenance methyltransferase because of its strong preference for hemimethylated DNA and direct association with newly replicated DNA (Leonhardt, H. *et al., Cell 71*:865-873 (1992)). To determine whether Dnmt3a and Dnmt3b polypeptides show any preference for hemimethylated DNA over unmethylated DNA, a comparison was done to examine the methylation rate of unmethylated versus hemimethylated oligonucleotides. Gel-purified double stranded oligonucleotides were incubated with protein extracts of Sf21 cells expressing Dnmt1, Dnmt3a, Dnmt3bl, Dnmt3b2 or NIH3T3 cell extract (unmethylated substrates (open circles), hemi-methylated substrates (half black diamonds) or completely methylated substrates

(closed squares)). While baculovirus-expressed *Dnmt1* polypeptide or 3T3 cell extract showed much higher activities when hemimethylated DNA was used as a substrate, Dnmt3a, Dnmt3b1 and Dnmt3b2 polypeptides showed no detectable preference for hemimethylated DNA (Fig. 8).